

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 27, 2001, 14:20:24 ; Search time 31.79 Seconds
(without alignments)
1046.951 Million cell updates/sec

Title: US-09-189-415A-2

Perfect score:

Sequence: 1 MPIGNLGNVNGNHLIPPAP.....GETAVSSVNAAPTGPVRFV 549

Scoring table: BLOSUM62

scoring curve: $\text{BESOM} \times \text{Gapop } 10.0$, $\text{Gapext } 0.5$

Searched: 412676 seqs, 60623988 residues

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0

Minimum DB seq	length: 0
Maximum DB seq	length: 2000000000

Post-processing: Minimum Match 0%

Post-processing: Minimum Match 0%
Maximum Match 100%

Database :

1:	/SIDS1/gcgdata/geneseq/geneseqp/AA1980.DAT:*
2:	/SIDS1/gcgdata/geneseq/geneseqp/AA1981.DAT:*
3:	/SIDS1/gcgdata/geneseq/geneseqp/AA1982.DAT:*
4:	/SIDS1/gcgdata/geneseq/geneseqp/AA1983.DAT:*
5:	/SIDS1/gcgdata/geneseq/geneseqp/AA1984.DAT:*
6:	/SIDS1/gcgdata/geneseq/geneseqp/AA1985.DAT:*
7:	/SIDS1/gcgdata/geneseq/geneseqp/AA1986.DAT:*
8:	/SIDS1/gcgdata/geneseq/geneseqp/AA1987.DAT:*
9:	/SIDS1/gcgdata/geneseq/geneseqp/AA1988.DAT:*
10:	/SIDS1/gcgdata/geneseq/geneseqp/AA1989.DAT:*
11:	/SIDS1/gcgdata/geneseq/geneseqp/AA1990.DAT:*
12:	/SIDS1/gcgdata/geneseq/geneseqp/AA1991.DAT:*
13:	/SIDS1/gcgdata/geneseq/geneseqp/AA1992.DAT:*
14:	/SIDS1/gcgdata/geneseq/geneseqp/AA1993.DAT:*
15:	/SIDS1/gcgdata/geneseq/geneseqp/AA1994.DAT:*
16:	/SIDS1/gcgdata/geneseq/geneseqp/AA1995.DAT:*
17:	/SIDS1/gcgdata/geneseq/geneseqp/AA1996.DAT:*
18:	/SIDS1/gcgdata/geneseq/geneseqp/AA1997.DAT:*
19:	/SIDS1/gcgdata/geneseq/geneseqp/AA1998.DAT:*
20:	/SIDS1/gcgdata/geneseq/geneseqp/AA1999.DAT:*
21:	/SIDS1/gcgdata/geneseq/geneseqp/AA2000.DAT:*
22:	/SIDS1/gcgdata/geneseq/geneseqp/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query %		Length	DB	ID	Description
		Match					
1	2798	99.9		549	20	AAV06220	EPEC E. coli trans
2	1492	53.3		559	20	AAV06221	EHEC E. coli trans
3	530	18.9		107	21	AAAB2076	Intimin C-terminal
4	165	5.9		30	20	AAV06213	EPEC E. coli trans
5	149	5.3		831	19	AAW68201	M. catarrhalis str
6	141	5.0		580	18	AAW10660	Endoglucoceramidas
7	139.5	5.0		2411	21	AAAB23860	Haemophilus influe
8	138.5	4.9		2353	17	AAAS93363	Haemophilus adhesi
9	136.5	4.9		3096	22	AAAB46770	FANCP3 protein fr
10	136.5	4.9		4630	18	AAW19629	Streptomyces venez
11	135	4.8		1532	21	AAAB40945	Human ORFX ORF709

12	133.5	4.8	1978	20	AAV27230	Amino acid sequenc
13	132.5	4.7	1026	18	AAV37490	Caulobacter cresce
14	132.5	4.7	1026	21	AAV44757	Caulobacter cresce
15	132.5	4.7	4630	15	AAV77177	S. venezuelae vep
16	131.5	4.7	1026	21	AAAR8993	rsaA S-lyase prote
17	130.5	4.7	1026	17	AAAR94014	Caulobacter S-lyase
18	129	4.6	1341	21	AAV95657	Human Acinus L pro
19	128.5	4.6	1536	18	AAW30293	Non-typeable Haemo
20	128	4.6	850	19	AAW62040	Protein isolated f
21	128	4.6	1237	18	AAW55640	H. pylori ORF 04ep
22	128	4.6	1237	20	AAV17187	H. pylori outer me
23	126	4.5	376	19	AAW81744	M. tuberculosis im
24	126	4.5	376	19	AAW64377	Mycobacterium tube
25	126	4.5	376	20	AAV39174	M. tuberculosis an
26	126	4.5	376	20	AAV39031	M. tuberculosis re
27	125.5	4.5	1354	21	AAV74857	Neisseria gonorrhoe
28	124.5	4.4	2514	21	AAV75097	Neisseria meningit
29	124	4.4	1180	21	AAAB01845	Haemophilus influe
30	124	4.4	1188	21	AAAB01844	Haemophilus influe
31	124	4.4	1536	15	AAAG3505	Haemophilus high m
32	124	4.4	1536	21	AAAB01846	Haemophilus influe
33	124	4.4	2541	21	AAAB41087	Human ORFX ORF851
34	123.5	4.4	3596	21	AAV87407	Bordetella pertuss
35	123.5	4.4	3547	11	AAAR05041	Filamentous haemag
36	123	4.4	741	9	AAAP80136	Neisseria iGA-prot
37	121	4.3	608	19	AAV97280	Arabidopsis phytoc
38	121	4.3	1638	20	AAV00138	Enterococcus faeca
39	121	4.3	1638	20	AAV00140	Enterococcus faeca
40	121	4.3	1638	20	AAV00142	Enterococcus faeca
41	121	4.3	2042	19	AAW56319	Haemophilus paraga
42	120.5	4.3	975	21	AAAB10763	Murine T protein f
43	120.5	4.3	2039	19	AAW36322	Haemophilus paraga
44	120	4.4	1048	18	AAW27277	Human cytomegalovi
45	119.5	4.3	1721	19	AAW52847	A. mediterranei ri

ALIGNMENTS

RESULT	1	
ID	AAAY06220	
XX	AAAY06220 standard; Protein; 549 AA.	
XX	AC	
XX	AAAY06220;	
XX	16-NOV-1999 (first entry)	
DT	XX	
XX	EPEC E. coli translocated intimin receptor (Tir).	
DE	XX	
XX	Tir; translocated intimin receptor; Hp90; enteropathogenic;	
KW	XX	
KW	EPEC; infection; diagnosis; vaccine.	
XX	XX	
OS	Escherichia coli.	
XX	XX	
PH	Key	Location/Qualifiers
FT	Domain	234..253
FT		/note= "putative transmembrane domain"
FT	Domain	364..386
FT		/note= "putative transmembrane domain"
FT	Misc-difference	180
FT		/note= "encoded by AAA"
FT	Misc-difference	314
FT		/note= "given as Xaa in the specification; Lys
FT		is deduced from the DNA sequence"
XX	XX	
PN	WO9924576-A1.	
XX	XX	
PD	20-MAY-1999.	
XX	XX	
PF	10-NOV-1998;	98WO-CA01042.
XX	XX	
PR	12-NOV-1997;	97US-0065130.
XX	XX	

XX	WO9924576-A1.
PN	
XX	20-MAY-1999.
PD	
XX	10-NOV-1998;
PF	
XX	12-NOV-1997;
PR	

QY 481 GYSVIQNFSGSPVTRGLIGTGTGQGIQSTYALLANSGLRLMGGLTSGGETAVSSVNA 540
Db 481 gysviqnfsgspvtrgligtgtgqgiqstyalansgrrlmggltsgetavssvnaa 540
QY 541 PTPGPVRFV 549
Db 541 ptpgpvrfrv 549
RESULT 2
AA06221
ID AAY06221 standard; Protein; 559 AA.
AC AAY06221;
DT 16-AUG-1999 (first entry)
DE EHEC E. coli translocated intimin receptor (Tir).
KW Tir; translocated intimin receptor; Hp90; enterohaemorrhagic;
KW EHEC; infection; diagnosis; vaccine.
OS Escherichia coli.
XX
FH Key Location/Qualifiers
FT Misc-difference 453 /note= "encoded by codon of 1 apparent nucleotide,
FT causing frameshift in the DNA sequence"
XX
PN WO9924576-A1.
XX 20-MAY-1999.
XX 10-NOV-1998; 98WO-CA01042.
XX 12-NOV-1997; 97US-0065130.
XX (UYBR-) UNIV BRITISH COLUMBIA.
XX Devinney R, Finlay BB, Kenny B, Stein M;
XX
XX WPI: 1999-337712/28.
XX N-PSDB; AAX58859.
XX
XX New translocated intimin receptor useful for treating infection by
XX enteropathogenic or enterohaemorrhagic Escherichia coli
XX Claim 7; Page 55-58; 91pp; English.
XX
XX The present sequence represents Tir, a novel translocated intimin
XX receptor (formerly termed Hp90) from an enterohaemorrhagic
XX Escherichia coli (EHEC) strain. The sequence was deduced from an
XX isolated tir polynucleotide (see AAX58859). Tir proteins are
XX secreted by attaching and effacing pathogens such as EHEC and EPEC
XX (see AAY06220) E. coli. The bacterial pathogens insert their own
XX receptors into mammalian cell surfaces, to which the pathogen then
XX adheres to trigger additional host signaling events and actin
XX nucleation. Diagnosis of disease caused by pathogenic E. coli can
XX be performed by use of antibodies that bind to Tir to detect the
XX protein or the use of nucleic acid probes for detection of nucleic
XX acids encoding Tir polypeptide. Isolated Tir nucleic acids, Tir
XX peptides, a recombinant method for producing recombinant Tir,
XX antibodies which bind to Tir, and a kit for the detection of
XX Tir-producing E. coli are provided. A method of immunising a host
XX with Tir to induce a protective immune response is also provided.
XX In addition, Tir fusion proteins can be used in attenuated E. coli
XX to induce a cell-mediated immune response to other polypeptides,
XX e.g. antigens. A method for screening for compounds which
XX interfere with the binding of bacterial pathogens to their
XX receptors is further provided.
XX Sequence 559 AA;

PA (UYBR-) UNIV BRITISH COLUMBIA.
XX Devinney R, Finlay BB, Kenny B, Stein M;
XX
XX WPI: 1999-337712/28.
XX N-PSDB; AAX58859.
XX
XX New translocated intimin receptor useful for treating infection by
XX enteropathogenic or enterohaemorrhagic Escherichia coli
XX Claim 6; Page 55-58; 91pp; English.
XX
XX The present sequence represents Tir, a novel translocated intimin
XX receptor (formerly termed Hp90) from an enteropathogenic
XX Escherichia coli (EPEC) strain. The sequence was deduced from an
XX isolated tir polynucleotide (see AAX58859). Tir proteins are
XX secreted by attaching and effacing pathogens such as EPEC and EHEC
XX (see AAY06221) E. coli. The bacterial pathogens insert their own
XX receptors into mammalian cell surfaces, to which the pathogen then
XX adheres to trigger additional host signaling events and actin
XX nucleation. Diagnosis of disease caused by pathogenic E. coli can
XX be performed by use of antibodies that bind to Tir to detect the
XX protein or the use of nucleic acid probes for detection of nucleic
XX acids encoding Tir polypeptide. Isolated Tir nucleic acids, Tir
XX peptides, a recombinant method for producing recombinant Tir,
XX antibodies which bind to Tir, and a kit for the detection of
XX Tir-producing E. coli are provided. A method of immunising a host
XX with Tir to induce a protective immune response is also provided.
XX In addition, Tir fusion proteins can be used in attenuated E. coli
XX to induce a cell-mediated immune response to other polypeptides,
XX e.g. antigens. A method for screening for compounds which
XX interfere with the binding of bacterial pathogens to their
XX receptors is further provided.
XX Sequence 549 AA;
Query Match 99.9%; Score 2798; DB 20; Length 549;
Best Local Similarity 99.8%; Pred. No. 1.1e-210;
Matches 548; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 MPIGNLNNVNGNHLIPAPPSPDGTGAARGTGHLISSTGALGSRSLFSPLRNSMADS 60
Db 1 mpignlgnvngnhiippepplpstgdgaargtgthllstgalgsrlfslprlnsmads 60
QY 61 VDSRDIPGLTPNPSLAATSTCLLGGFVLHDKGDLILNTQIGPSAFRVEVQADGTH 120
Db 61 vdsrdipgltpnpsrlaaatsetcll9gfevindhk9pdlilntqigpsafrevevqadgth 120
QY 121 AATGKNGLEVSTLSPQWSSLSQIDTEGKNRFTVGTGGSGGHPMTVVASDIAEARTR 180
Db 121 aatgknglevstlspqewslqsidteqknrfvftggrgsgghpmvtvasadiaeartr 180
QY 181 ILAKLPDNNHGRQPKDVTNRVGVSGASIDGVVSETHSTTNSVRSRDPKFWVSVA 240
Db 181 ilaklpdnhgrqpkdvtrsvgvsgasidgvvsethsttntssvrsrdbkfwvsvga 240
QY 241 IAAGLAGLAATGAOALATPEPDPTTTDPOAANAESATKDOLTOFAFNKPNQKYN 300
Db 241 iaaglaglaatgaalatlpepdpttttdpoaanaesatkdltoafafnkpnqkyn 300
QY 301 IDANGNAIPSGELXDDIVEIOIAQAKAEVAPQAVESNAQAQRYEDOHARRQEELQ 360
Db 301 idangnaipsgelkddiveioiaqakeavearqavessnaqaqrqedgharrqeelq 360
QY 361 SSGIGVLSALLIVAGGIGAVGTALHRRNPQAEOTTTTHTVVQQQTGGIPQHKVALL 420
Db 361 ssgigvllsallivaggigavgtalhrnpqaeqtthtthlvvqqqtggipqhkvall 420
QY 421 PQERRRFRDRDQSGSVASPTHWSDDSSSEVVNPNYAEGGARNLSAHPQEEHYDEVAADP 480
Db 421 pqerrfrdrdsgsvasthwsdsssevvnpnyaevggarnlsahpqeehydevaadp 480

```
Query Match      53.3%; Score 1492; DB 20; Length 559;
Best Local Similarity 54.7%; Pred. No. 1.7e-108;
Matches 322; Conservative 64; Mismatches 133; Indels 70; Gaps 12;

QY 1 MPIGNLNNVNGNHLIPAPPLPSQTDGAARGGTGHLISSTGALGSRSLFSPLRNSMADS 60
DB 1 mpignlghnpvnnsippapplpstqda--ggrgqlinstgplgsralftprynsmads 58

QY 61 VDSR--DIPGLTNPSSRLAATSETCLLGGFEVLHDKGLDILNTOIGPSAPFRVQADG 118
DB 59 gdnrasdvplgvnmpmlaa--seitlndgfevlhdkpldlnrqigsvfrvetqdg 116

QY 119 TAAIGCKNGLEVSFTLSPQESSLQSDTECKNRFVFTGGRGGSGHPMVTVASDIAEAR 178
DB 117 khlavgringvetsvlsdqeyarqsiqidegkdvfvfgrggghamvtvasditear 176

QY 179 TRILAKLPDNNHGRQPKDVTFRSVGVGSASGI-----DDGV--VSEPHTTSTNSSVRS 230
DB 177 qrillelepkgtg-----eskgageskgvgelresnsgaenttetqtststslrs 227

QY 231 DKFWVSGAIAAGLAGLAATGIAQALATPEPDDPTTDPQANAABESATKDQLTQEA 290
DB 228 dplkwlalgtvatgliglaatgvlqalaitpepdsptttdpdaasatetatrdqtklea 287

QY 291 FKPNQKVNIDANGNAIPSGELXDDIVQIAQAKEAGEVARQQAQVESNAQAQORYEDQ 350
DB 288 fgnpdnqknidelnaiapsgvylkdvvanieeqkaageekqqaiaennaqakkydeq 347

QY 351 HARQELQLSSGIGYGLSSALIVAGGAGVTTALHRRNPQAEQTTTTT-----HT 403
DB 348 qakrqeelkvssgagylsgalllggigvavtaalhrknqveqtttttttttsart 407

QY 404 VVQQQTGGTPQHKVALMPQERRRDSRDSQCSVASHTHWSDSSEVVNPYAEVGG---AR 460
DB 408 venkpanntpcaggnvtdpgsedtmesrrssmaststffdtss-----lpgpccrir 458

QY 461 NLSIAHQ-----PEEHYDEVAADPGSVIQNFSGSGPVTGRLIG 500
DB 459 mmlmklhrcmirrcrlllrlfrwigiqisvvytiqhpp-----rdtdng---arlig 510

QY 501 TPGQGTQSTYALLANSGLGURLGMGGJTGSETAVSSVNAAPTGPVRFV 549
DB 511 npsagiqtayarlalsgglrhdmgltggsnsavntsnpppggshrfv 559

RESULT 3
AAB20576
ID AAB20576 standard; protein; 107 AA.
XX
AC AAB20576;
XX
DT 08-DEC-2000 (first entry)
XX
DE Intimin C-terminal Tir binding domain amino acid sequence.
XX
KW Intimin; Tir binding domain; inhibitor; intimin adhesion; screening;
KW Tir-independent eukaryotic cell binding activity; bacterial infection;
KW diarrhoea; antibacterial.
XX
OS Unidentified.
XX
PN WO200045173-A1.
XX
PD 03-AUG-2000.
XX
PF 31-JAN-2000; 2000WO-GB00254.
XX
PR 29-JAN-1999; 99GB-0001897.
XX
PA (IMCO-) IMPERIAL COLLEGE INNOVATIONS LTD.
XX
PI Frankel GM, Matthews SJ, Hale CB, Dougan G;
```

```
XX WPI; 2000-499357/44.
XX
PT Screening for inhibitors of intimin binding to eukaryotic cells, for
PT use in diagnosing, preventing and treating bacterial infections,
PT especially Escherichia coli O157:H7 -
XX
PS Claim 8; Page 76; 96pp; English.
XX

CC The present invention describes a method of screening for an inhibitor
CC of intimin binding to eukaryotic cells. The method comprises exposing an
CC intimin polypeptide having a Tir-independent cell binding activity to
CC test agents, and obtaining an inhibitor based on its ability to bind the
CC polypeptide. The inhibitors are used in the prevention, treatment and/or
CC diagnosis of bacterial infections, preferably by enteropathic and/or
CC enterohaemorrhagic Escherichia coli, Shiga toxinigenic E. coli, Hafnia
CC alvei or Citrobacter freundii, or especially E. coli O157:H7. The
CC effacement on intestinal epithelial cells. The inhibitors can be used
CC to produce food supplements or additives, especially where the food is
CC a milk substitute. The method can be used to sort cells based on their
CC ability to bind to a Tir independent cell binding domain of an intimin
CC polypeptide. Polypeptides having Tir-independent intimin binding
CC activity can be used to produce a vaccine against a bacterial disease.
CC The present sequence represents a specifically claimed intimin C-terminal
CC Tir binding domain amino acid sequence, for use in the method of the
CC present invention.
XX
SQ Sequence 107 AA;

Query Match      18.9%; Score 530; DB 21; Length 107;
Best Local Similarity 99.1%; Pred. No. 3.1e-34;
Matches 106; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 255 QALALTPEPDDPTTDPQANAABESATKDQLTQEAFAKNPENQKVNIDANGNAIPSGELX 314
DB 1 qalaltpepddptttdpqaanaaesatkdtqeaafknpenqkvnidangnaipsgelk 60

QY 315 DDIVQIAQAKEAGEVARQQAQVESNAQAQORYEDQHAHQEELQLS 361
DB 61 ddiveqiaqakeagevarqqaqvesnaqagqrqedgharreeelqls 107

RESULT 4
AAY06213
ID AAY06213 standard; Peptide; 30 AA.
XX
AC AAY06213;
XX
DT 16-AUG-1999 (first entry)
XX
DE EPEC E. coli translocated intimin receptor N-terminal peptide.
XX
KW Tir; translocated intimin receptor; Hp90; enteropathogenic;
KW EPEC; infection; diagnosis; vaccine.
XX
OS Escherichia coli.
XX
PN WO9924576-A1.
XX
PD 20-MAY-1999.
XX
PF 10-NOV-1998; 98WO-CA01042.
XX
PR 12-NOV-1997; 97US-0065130.
XX
PA (UYBR-) UNIV BRITISH COLUMBIA.
XX
PI Devinney R, Finlay BB, Kenny B, Stein M;
XX
DR WPI; 1999-337712/28.
XX
```

PT New translocated intimin receptor useful for treating infection by
PT enteropathogenic or enterohemorrhagic Escherichia coli
XX
XX
XX Example 1; Page 37; 91pp; English.
XX
XX The present sequence represents the N-terminal sequence of Tir (see
CC also AAY06220), a novel translocated intimin receptor from an
CC enteropathogenic Escherichia coli (EPEC) strain. The 78 kDa EPEC
CC protein is secreted by the bacterial pathogen. Diagnosis of
CC disease caused by pathogenic E. coli can be performed by use of
CC antibodies that bind to Tir to detect the protein, or the use of
CC nucleic acid probes for detection of nucleic acids encoding Tir. A
CC kit for the detection of Tir-producing E. coli is provided. Also
CC provided are a method of immunising a host with Tir to induce a
CC protective immune response, and a method for screening for
CC compounds which interfere with the binding of bacterial pathogens
CC to their receptors.
XX
XX
SQ Sequence 30 AA;

Query Match 5.9%; Score 165; DB 20; Length 30;
Best Local Similarity 100.0%; Pred. No. 1.9e-06;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2 PIGNLGNVNGNHLIPAPPPLPSQTGGAAR 31
DB 1 pignlgnvngnhlipappplpsqtdgaar 30

RESULT 5
AAW68201
ID AAW68201 standard; Protein; 831 AA.
XX:
AC AAW68201;
XX
XX 07-OCT-1998 (first entry)
XX M. catarrhalis strain O35E UspA1 antigen.
XX Moraxella catarrhalis; UspA1; UspA2; antigen; genetic vaccination;
KW vaccine; otitis media; sinusitis; lower respiratory tract infection;
KW immunity enhancer; immunoassay reagent.
XX
OS Moraxella catarrhalis.
XX
XX WO9828333-A2.
XX
XX 02-JUL-1998.
XX
XX 19-DEC-1997; 97WO-US23930.
XX
XX 20-DEC-1996; 96US-0033598.
XX (TEXA) UNIV TEXAS SYSTEM.
XX Aebi C, Cope LD, Fiske MJ, Fredenburg R, Hansen EJ;
PI Maciver I;
XX
XX WPI; 1998-377595/32.
DR N-PSDB; AAV41341.
XX
XX New peptide(s) containing the core epitope of Moraxella catarrhalis
PT Usp proteins - useful in, e.g. vaccines to prevent or treat M.
PT catarrhalis infection, and antibodies for passive immunisation
XX
XX Claim 22; Pages 133-136; 237pp; English.
XX This represents a UspA1 antigen of Moraxella catarrhalis strain O35E.
CC Nucleic acid sequences encoding the UspA1 and A2 antigens of
CC M. catarrhalis isolates O35E, O46E, TTA24 and TTA37 can be used in
CC genetic vaccination. An antigenic composition or vaccine containing
CC antigenic peptides from UspA1 or UspA2 antigens are used to induce an

CC immune response in mammals against M. catarrhalis and can be used to
CC treat infections such as otitis media, sinusitis, lower respiratory
CC tract infections. They can also be used as immunity enhancers for other
CC bacterial, parasitic or viral antigens, to raise antibodies and as
CC immunoassay reagents for detecting specific antibodies. The antibodies
CC are useful for passive immunisation and as immunoassay reagents.
CC detection of the epitopic core sequence, by immunoassay or by PCR, is
CC used to diagnose infection. The usp antigens encoding nucleic acid
CC sequences are also used to produce recombinant proteins and for screening
CC for potential anti-M. catarrhalis agents, while their fragments are
CC useful as diagnostic probes or primers or to isolate variant sequences.
XX
XX Sequence 831 AA;

Query Match 5.3%; Score 149; DB 19; Length 831;
Best Local Similarity 19.2%; Pred. No. 0.0033;
Matches 101; Conservative 73; Mismatches 193; Indels 160; Gaps 21;
QY 74 SRLAAATSETCLLGFEVLHDKGPLDLNTQIGP---SAFRVEVQADGTHAIGEKNGLE 130
DB 226 sgttgvtstnsvllgneta--gkqattvknaevgglsaltgfageskaengvavsgsegr 283
QY 131 VSYTLSPQEWSSLOSIDTRGKNRFVFTGGGGSGHPMTVAS----DIAEARTRILAKLD 186
DB 284 qivnvga---ggisdstdavn-----gsqhalatvvdnqydivnnradilnnqd 332
QY 187 P-----DNHGGROPKVDVT-----RSVGVGS----- 207
DB 333 dikhqkevkgldnevgelsrdinslhvtdnqgddikelkrgvkeldnevgvlrsdins 392
QY 208 -----ASGIDDVVSETHSTTNSVRSOPKFWVSYGATAAGLAGLAATGIAQALALTP 262
DB 393 lhadvadnqddiaknkadikglmkvkeidk---evgvisrdigsl----- 435
QY 263 PDDPTTTDPQAAANAESATKDQLTQEAENPE-----NOKVNIDAN----- 304
DB 436 hddvatnqadiaknqadiktlennveeellnlsgrlldqkadidnnlnniyelaqqdqh 495
QY 305 -----GNAIPSG--ELXDDIVEQIAQQAKEAGEVARQAQVESNAQAQOQRYEDQHARQ 355
DB 496 ssdiktiknnveeglldslgrlldqkadiaknqadiagnqtdqldlaaynelqdyqyqkq 555
QY 356 EE-----LQLSSGIG--YCL-----SSALIVAGGIGAGVTTAL 386
DB 556 teaidalnassentqniaknqadianniniyelaqqdqhssdiktakvsaantdri 615
QY 387 HRRNQPAE--QTTTTHTVVQQQTGQIPQHKVALMPQRRRFRSDRRDSQGSVASTHWS 443
DB 616 aknkaeadasfettiknqntlieqgealveqnka--lnqlegfaahadiq----- 664
QY 444 DSSSEVNPYAEVCGARNSLSAHQPEEHYDEVAADPGYSVIONFSG 490
DB 665 --dkqilqnqaditntkai-----eqnairtva--ngfelekknag 702

RESULT 6
AAW10660
ID AAW10660 standard; Protein; 580 AA.
XX
XX AAW10660;
XX
XX 20-JUL-1997 (first entry)
DE Endoglycoceramidase activator II.
XX
XX Endoglycoceramidase; activator; glycolipid; enzyme.
XX Rhodococcus spp. M-777.
XX
XX EP759470-A2.
XX
XX 26-FEB-1997.
PD

XX 28-JUN-1996; 96EP-0110513.
 XX 29-JUN-1995; 95JP-0188466.
 XX (TAKI) TAKARA SHUZO CO LTD.
 XX Ito M, Izu H, Izumi Y, Kato I, Kurome Y, Sano M;
 XX WPI; 1997-147519/14.
 DR N-PSDB; AAT61050.
 XX
 XX DNA encoding endo:glyco:cer:amidase activator polypeptide - for
 PT prodn. of recombinant polypeptide, useful in glyco:lipid analysis
 XX
 XX Claim 1; Page 13-15; 30pp; English.
 PS
 XX DNA encoding endoglycoceramidase activator polypeptide can be used
 CC for the prodn. of the recombinant polypeptide, which can be used
 CC to research the intracellular function of glycolipids, because
 CC it shifts the optimum pH for endoglycoceramidase II towards
 CC neutral, allowing the latter to be used to hydrolyse glycolipids
 CC even at pH 7.5.
 XX
 XX Sequence 580 AA;

Query Match 5.0%; Score 141; DB 18; Length 580;
 Best Local Similarity 23.9%; Pred. No. 0.0085;
 Matches 84; Conservative 45; Mismatches 126; Indels 96; Gaps 16;
 QY 4 GNLGNVNGNHLIPP-----APPLPSQTDGAARGGTGHLISSTGALGSRSLFSP 53
 DB 215 gqfndanfllflprasapivgtwaptqcpdrtaa-----gplnagagplatiqi---l 267
 QY: 54 RNSMADSVDSRDIPLPNPSR--LAAATSEFCLLGFEVLHD---KGPLDILNTQIGP 107
 DB 268 rqaava-tvsyldgpsavtnggftinatvptpdsqgvqftrdgedgapvdlvn---gk 323
 QY 108 SAFRVEVQADGTHAAGEKNGLEV-----SVILSPQEWSSLOSIDTEGKNRFVFGG 159
 DB 324 asltsgldtdgdyaeyakflgaefnfpssaaktvtstgdiqttsvt----- 371
 QY 160 RGGSGHPMVTASDIAEATRLAKLPDNGHGRQPKVDTRSVGVSGASGIDGQVWSET 219
 DB 372 ---gpdh-----dayrdpvnltakvepgvsggtvafevd--gtpvgtdavmdgaaalp 421
 QY 220 HTSTNSSVRSDPKFVSVGAIAAGLAGLAATGIAQALAL-----TPEPDDPTTTTDP 271
 DB 422 hfttngthr-----viarysg--aegispsvslqypsvteapaadvattitv 468
 QY 272 DQANNAESATKDQLTQEAFAKNPENQK-----VNIDANGNA 307
 DB 469 dplaatags---pvtlariqpdaargvtvqfklgdvllgpgvrvdangva 516

RESULT 7
 AAB23860
 ID AAB23860 standard; Protein; 2411 AA.
 XX
 AC AAB23860;
 XX
 DT 17-JAN-2001 (first entry)
 XX
 XX Haemophilus influenzae adhesin (Hia) protein from type c strain API.
 DE
 XX Hia; adhesin; Haemophilus influenzae adhesin; NTHI; infection; vaccine;
 KW non-typeable Haemophilus influenzae; antiinflammatory; auditory;
 KW antibacterial; meningitis; epiglottitis; septicaemia; otitis media;
 KW diagnosis; immunogenic; antigen.
 XX
 OS Haemophilus influenzae.
 XX

PN WO200055191-A2.
 XX
 PD 21-SEP-2000.
 XX
 PF 16-MAR-2000; 2000WO-CA00289.
 XX
 PR 16-MAR-1999; 99US-0268347.
 XX
 PA (CONN-) CONNAUGHT LAB LTD.
 XX
 PI Loosmore SM, Yang Y, Klein MH;
 XX
 DR WPI; 2000-618897/59.
 DR N-PSDB; AAA92499.
 XX
 PT Novel nucleic acid encoding Hemophilus influenzae adhesin protein, for
 PT use as antigens and vaccines and for treating Hemophilus influenzae
 PT infection -
 XX
 PS Claim 1; Fig 24; 275pp; English.
 CC
 CC The present sequence represents a Haemophilus influenzae adhesin
 CC (Hia) protein from the type c Haemophilus influenzae strain API.
 CC Hia genes and proteins have antiinflammatory, auditory and antibacterial
 CC activities, and can be used in the production of a vaccine. An
 CC immunogenic composition comprising an Hia gene, a polypeptide encoded
 CC by an Hia gene, or a recombinant Hia polypeptide is useful for inducing
 CC protection against disease caused by Haemophilus strains in a
 CC susceptible host, preferably a human. An Hia protein is useful as an
 CC antigen, in immunogenic preparations including vaccines, as a carrier
 CC for other immunogens, and in the generation of diagnostic reagents. Hia
 CC is useful for treating diseases caused by the infection of Haemophilus
 CC influenzae such as meningitis, epiglottitis, septicaemia and otitis
 CC media. Recombinant production of Hia favours high recovery of the
 CC protein compared to the low recovery of native protein from Haemophilus
 CC influenzae species. A truncated protein has a significantly higher
 CC amount of recovery than a full-length protein.
 XX
 XX Sequence 2411 AA;

Query Match 5.0%; Score 139.5; DB 21; Length 2411;
 Best Local Similarity 21.5%; Pred. No. 0.079;
 Matches 125; Conservative 83; Mismatches 228; Indels 145; Gaps 28;
 QY 6 LGNNVNGNH---LIPPAPPLPSQTDGAARG---GTGHLISST--GALGSRSLFSPLRNS 56
 DB 1161 lqgtangrndtgvinkdglttitlangaaagtdaangntisvtkdgisagnkeintvksa 1220
 QY 57 MADSVDSRDIPLPNPSRLAAATSETCLLGFEVLHDKGPLDILNTQIGPSAFRVEVOA 116
 DB 1221 lkytdtqntag-atqp---aantaeva-----kqdlvdltpatgaagnadaka 1267
 QY 117 -DGHAAIGEKNGL-----EVSVTLSPOEW-----SSLSQSIDTEGKNRFVFTGGGGSG 164
 DB 1268 pdttaatvgdlrglswlsakktadtdqkfehaavknevefvgkngatvsaktddng 1327
 QY 165 HPMTVVASDIAEAR-----TRILAKLPDNGHGRQPKVDTRSVGVSGASGIDGQV 216
 DB 1328 khtvti--dvaekvgdglektdgkikikvndtgnnllttda-tkgasvakgefnavt 1384
 QY 217 SEHTST-TNSSVR-----SDPKFVSVGAIAAGLAGLAATGTAQALALT 260
 DB 1385 tdataggtanergkvvkvngngatataetdkkvtatvgdvakaind-aatfv-----k 1437
 QY 261 PEPDDPTTTDPQANNAESATK--DQLTQEAFAKNPENQKVNID-----ANGNAIPS 310
 DB 1438 venddsatiddspdtgdandalkagdtltlkagk---nlkvrdgknitfalndlsvks 1494
 QY 311 GELXDDIVEQIAQQAKEAGEVARQQAQRYEDQHARRQEEQLSSGIGYGLSS 370
 DB 1495 atvskdl-----slgtngknknitsdtkglnfakdsktgddanhl-ngiastltd 1544

QY 371 ALIVAGGIGAGVTTALHRRNQPAEQTTTTHTTVVQQQTGGIPQHKVLMPOERRRFSR 430
Db 1545 tl-----lmsgattnl-----gg-----ngitdnekkraasv 1571
QY 431 RDSQGSVASTHWSDSSESVNPAEYGGARNLSAHOPEEHYDEVAADPGYSVIONFSG 490
Db 1572 kd-----vlnagvnrvgkpa-sannqveni-dfvt---ydtvdfvsg 1610
QY 491 SGPVTRGLTPOGGIQSTYVALLANGGLRGLMGGLTSGGE 531
Db 1611 dkdttsvtveskdngkrtvkgaktsvikdhngkiftgke 1651

RESULT 8
AAR99393
ID AAR99393 standard; Protein; 2353 AA.
XX AC AAR99393;
XX DT 15-JAN-1997 (first entry)
XX DE Haemophilus adhesion protein HA2.
XX KW Haemophilus adhesion protein; HA2; hsf protein; vaccine.
XX OS Haemophilus influenzae type b strain C54.
XX PN WO9630519-A1.
XX PD 03-OCT-1996.
XX PF 22-MAR-1996; 96WO-USO4031.
XX PR 24-MAR-1995; 95US-0409995.
XX PA (UYSL-) UNIV ST LOUIS.
XX PA (UNIW) UNIV WASHINGTON.
XX PI Barenkamp SJ, St Geme JW;
XX WPI; 1996-455364/45.
DR N-PSDB; AAT41476.
XX Recombinant Haemophilus adhesion proteins HA1 and HA2 - for use in
XX vaccines against H. influenzae infection.
XX Claim 5; Page 66-73; 120pp; English.
XX Haemophilus adhesion protein HA2 (AAR99393) is associated with the
XX formation of surface fibrils involved in adhesion to various host
XX cells; it is also referred to hsf (Haemophilus surface fibrils).
XX Its amino acid sequence was deduced from a genomic DNA clone
XX (AAT41476) derived from Haemophilus influenzae type b strain C65.
XX Large quantities of recombinant HA2 can be produced in transformed
XX prokaryotic or eukaryotic host cells, for use in vaccines against
XX H. influenzae infection.
XX Sequence 2353 AA;
SQ

Query Match 4.9%; Score 138.5; DB 17; Length 2353;
Best Local Similarity 22.0%; Pred. No. 0.091;
Matches 124; Conservative 62; Mismatches 207; Indels 171; Gaps 26;
QY 6 LGNNVGNHLIPPAPLPSTQDGAARGGTGHLISSTGALGSRSLFSLRNSMADSVDSRD 65
Db 1163 lqgtang-----indgtvinkdg-----litiangaagtdaen 1198
QY 66 IGPLTPNPSRLAAATSETCLLGGFEVLHDKGPL-----DILNTQIGSAPRVEVQADGTHA 121
Db 1199 -----gntisvtdkgisagnkeitnvksalktykdtqnt-----adetqdkfha 1243
QY 122 AIGKKNGLVSVTLSPQEWSSLSQSIDTEGKNRFVFTGGRGSGHPMVTVASDIAEAR--- 178

Db 1244 av--knanevefv-----gknagatvsaktdnngkhtvti--dvaekavgd 1284
QY 179 -----TRILAKLDPNHGGRQPKVDTRSVGSGASGIDGQVSVSETHST-TNSSVR--- 229
Db 1285 glekddgkiklkvndtdgnlltvda-tkgasvakgefnavttdattaggtannergkv 1343
QY 230 -----SDPKFWVSVGAIAAGLAGLAATGTAQALALTPPEPDDPTTDPQOANA 277
Db 1344 vvksgngatatetdkkvatvgvakaind-aatfv-----kvenddsatiddsptddg 1396
QY 278 AESATK--DQLTQEFKPNENKVNI-----ANGNAIPSGELXDDIVBOIAQOAKE 327
Db 1397 andaikagdtlilkagk---nlkvkrdgnkniitfalandslvsksatvsdkl-----s 1444
QY 328 AGEVARQOAVESNAQAQRYEQHARRQEELQLSSGIGYLSALIVAGGIGAGVTTALH 387
Db 1445 lgtngknvitsdtkgnlfakdsktgddanihl-ngiastltdtl-----lmsgattnl- 1497
QY 388 RRNPAPAEQTTTTHTTVVQQQTGGIPQHKVLMPOERRRFSRDRDSQGSVASTHWSDS 447
Db 1498 -----gg-----ngitdnekkraasvkd----- 1515
QY 448 EYVNPYAEVGGARNLSAHOPEEHYDEVAADPGYSVIONFSGSPVTRGLTGPQGGIQ 507
Db 1516 -vlnagvnrvgkpa-sannqveni-dfvt---ydtvdfvsgkdttsvtveskdngkr 1569
QY 508 STYALLANGSGLRLGMLTSGGE 531
Db 1570 tevkgaktsvikdhngkiftgke 1593

RESULT 9
AAB46770
ID AAB46770 standard; Protein; 3096 AA.
XX AC AAB46770;
XX DT 19-APR-2001 (first entry)
XX DE FANCIP3 protein fragment SEQ ID NO 10.
XX KW FANCIP2; FANCIP3; antitumor; signal transduction cascade; immunogen;
XX Fanconi anemia complementation group A protein; DNA-repair defect;
XX cell-cycle disorder; cytopenia; tumorigenesis; tumor progression;
XX gene therapy.
XX OS Unidentified.
XX PN WO200100822-A2.
XX PD 04-JAN-2001.
XX PF 26-JUN-2000; 2000WO-EF05878.
XX PR 29-JUN-1999; 99DE-1029887.
XX PA (MULT-) MULTIGENE BIOTECH GMBH.
XX PI Gross HJ, Reuter T, Hanenberg H, Herterich S, Wagner M;
XX WPI; 2001-102892/11.
XX New nucleic acid encoding Fanconi anemia protein interacting proteins,
XX useful e.g. for diagnosis and treatment of DNA-repair defects -
XX Claim 1; Page 49-58; 60pp; German.
XX This invention describes a novel nucleic acid (I) which has antitumor
XX activity. (I) encode proteins that interact with the Fanconi anemia
XX protein of complementation group A, so may be part of the complex or
XX signal transduction cascade which causes the defect responsible for
XX Fanconi anemia. Polypeptides (II) encoded by (I) are used as immunogens

CC for preparation of antibodies (Ab), also for identifying specific
 CC modulators (A). Any of (I) (or vectors or cells that contain (I)), (II),
 CC Ab and (A) are useful for: (a) diagnosis of diseases (or predisposition
 CC to them) associated with DNA-repair defects, cell-cycle disorders,
 CC cytopenia, tumorigenesis and tumor progression; and (b) for treatment
 CC (including by gene therapy) or prevention of these diseases.
 XX
 XX Sequence 3096 AA;

Query Match 4.9%; Score 136.5; DB 22; Length 3096;
 Best Local Similarity 21.0%; Pred. No. 0.19;
 Matches 144; Conservative 78; Mismatches 245; Indels 219; Gaps 33;

QY 18 PAPPLPSQTDGARGGCTHLISSTGALGSRSLFSLRNSMADSVSRDIP--GLTNPSPR 75
 Db 1916 pasplys---gialagl-----nlelysl-nglaslnysmetpheglnasple 1960
 QY 76 LAAATSETC-----LLG-----GPEVLHDKGP--LDILNTQ-----IG 106
 Db 1961 leglyltyrlyserleglylgyshisargargvallysaspalaeasnthrthrgla 2020
 QY 107 PSAFRVEVQADGTHAAGKNGLEV--SVTLSPQEWSSLSQSDT-----E 149
 Db 2021 snlyslleaspalatyrthrglnleaseraspelysargmetileserlyleaglala 2080
 QY 150 GKNRFVFTGGR-----GGSGHPWTVASDIARPRILAKLPDNDHG-----G 192
 Db 2081 glnvallysglnvalghisglnasmetleserlearghisasnerargilehsvala 2140
 QY 193 ROPKQDVTRSVGSGASGIDGWSETH---TSTNSSVRSRDPKFWVSGAAGLAGLA 249
 Db 2141 rpr--serargalaa-----snthrealeathrerasp---valserargarly 2187
 QY 250 AT-----GIAQALALT-----PEPD-----PTTDPDQAANAAE 279
 Db 2188 strpleileprglyalagityrserilephethrlyglnrleaspt----hrglnasp 2243
 QY 280 SATKQOLTQEAFFKNQKVNIDANGNAIPSGELXDDIVEQIAQAKAGEVARQOAVES 339
 Db 2244 s---erasnvalaspasnglnle--glgthrcysserleglyhisargserprlegly 2298
 QY 340 NAAQAQRYE-----DQHARRQBELQLSSGIGYGLLSALIVAGGIGAGVTTALH 387
 Db 2299 saspserprglyseraserthrerlelellylsglnargl-----th 2351
 QY 388 RNQPAEQTTTTHTVVQQQG-----GIPQHKVALMPQERRRFS 428
 Db 2352 r-----serasphrilemetargalalelylsglleaspglylylsi-lephelysas 2406
 QY 429 DRDS-----QGSVASTHWSDSSEVW---NPYAEVGGARNLSAHQPEE 470
 Db 2407 ntrpglythrlnthrglylsglasphraserasnlleasnnp-----rargglnthrgit 2461
 QY 471 HIYDEVAADPGYSVQNFSGSGPVTGRIGTPGQGIQSTYALLANSGLRLMGGL----- 526
 Db 2462 hrservalasnalaserargserprgll-----yscysalaglnln-argglnlyls 2512
 QY 527 -----TSGETAVSSVNAAPTGPVR 547
 Db 2513 argleasnseralaserargser 2538

RESULT 10
 ID AAW19629
 XX AAW19629 standard; Protein; 4630 AA.
 AC AAW19629;
 XX AAW19629;
 XX 01-SEP-1997 (first entry)
 DE Streptomyces venezuelae polyketide synthase.

KW Polyketide synthase; polyhydroxyalkanoate monomer synthase;
 KW polyhydroxybutyrate; biodegradable polymer; vep gene;
 KW metabolic engineering.
 XX Streptomyces venezuelae.
 OS WO97272711-A1.
 PN 26-JUN-1997.
 PD 18-DEC-1996; 96WO-US20119.
 PF 19-DEC-1995; 95US-0008847.
 PR (MINU) UNIV MINNESOTA.
 PA Sherman DH, Williams MD, Xue Y;
 PI WPI; 1997-341701/31.
 DR N-PSDB; AAT68715.
 DR Expression cassettes for production of polyhydroxyalkanoate(s) -
 PT provide wide range of biodegradable polymers for medical or
 PT industrial use
 XX Claim 55; Fig 23; 91pp; English.
 PS 3 Polypeptide sequences (AAW19629-30 and AAW00918) can be deduced from
 CC the vep ORF1 polyketide synthase (PKS) gene cluster (AAT68715) of
 CC Streptomyces venezuelae. The sequence data indicate that the PKS
 CC gene cluster encodes a polyene of 12 carbons. The vep gene cluster
 CC contains 5 PKS modules, plus a 5' loading module and a 3' end
 CC domain. Each of the sequenced modules includes a keto-ACP, an
 CC acyltransferase, a dehydratase, a keto-reductase and an acyl carrier
 CC protein domain. A novel expression cassette encoding the first
 CC module from the vep gene cluster and module 7 from the Streptomyces
 CC tyLP gene cluster has polyhydroxyalkanoate (PHA) monomer synthase
 CC activity and can be used for PHA prodn. in host (esp. insect) cells
 CC for use as a biodegradable polymer.
 XX Sequence 4630 AA;

Query Match 4.9%; Score 136.5; DB 18; Length 4630;
 Best Local Similarity 22.2%; Pred. No. 0.33;
 Matches 125; Conservative 49; Mismatches 203; Indels 187; Gaps 27;

QY 23 PSQTDGAARGGCTHLISST--GALGSRSLFSLRNSMADSVSD---SRDIPGLTPNPSRLA 77
 Db 1156 prmhdg-aegvegghlltgttspvmgriayqilgtgavtdtaccssivalhavrslr 1214
 QY 78 AATSETCLLGFEVLHDKGPLDILNTQIGPSA-----FRVEVQADGTHAAGKNGLEVSV 133
 Db 1215 qgesslallagatvmstpgmfefsrgrlaadgrskafdsadgtsaegv--gllvve 1272
 QY 134 TLSPOEWSSLSQSDIDTEGKNRFVTGGRGSGHPMTVASDIA-----EAR 178
 Db 1273 rlsdae-----rnhpvlavirgsavngdngasngltapngpsq 1310
 QY 179 TRILAKLPDNDHGRQPKDWD--TRSVGVSASGIDGVVSETHSTTN-----S 226
 Db 1311 qrvirqalad--agltpadvdaveahgtgrlg--dpieaeailgtgydrgegaplqig 1366
 QY 227 SVSRDPKFWVSGATAAGLAGLAATGIAQALALTPBP---DDPTTDDPOAANAASATK 283
 Db 1367 slksn---iahaqaaagvgglikmvlamrhgvlprtlhvdprptr-vdweagggvellt 1421
 QY 284 DQLTQEAFFKNPENQKVNIDANGNAIPSGELXDDIVEQIAQAKAGEVARQOAVESNAQA 343
 Db 1422 er-----ewpetgr-----prraaisffgtsnahlivveqapeage----- 1458
 QY 344 QORYEDQHARRQBELQLSSGIGYGLLSALIVAGGIGVTTALHRRNQPAE--QTTTTTT 401

Db 1459 -----aavttapeageageadtattt 1482

Qy 402 HTVVQQTGGIPQ-----HKVALMPQ--ERRRSDRDSOGSVASTHWSDSSE 448

Db 1483 paav-----gvpeprapvvvsardaalaarqavrlrtfldgr----- 1520

Qy 449 VVNPYARVGGARNSLAHQPEEH-----IYDEVAADPGYSVIONFSGSG-PVTGRLIG 500

Db 1521 ---pdvctvadglsraafefehkaaltatrdella--gldal-----grgeatglvtg 1571

Qy 501 TPGGQIQSTYALLANSGLRLGMG 524

Db 1572 eparaartaf-lftgggaqrvang 1594

RESULT 11

AAB40945

ID AAB40945 standard; Protein; 1532 AA.

AC AAB40945;

XX 08-FEB-2001 (first entry)

Human OREF709 polypeptide sequence SEQ ID NO:1418.

XX Human; open reading frame; OREFX; detection; cytostatic; hepatotropic;

KW vulnary; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;

KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;

KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;

KW hypotensive; dermatological; immunosuppressive; antiinflammatory;

KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;

KW antianaemic; gene therapy; cancer; proliferative disorder; hypertension;

KW neurodegenerative disorder; osteoarthritis; graft vs host disease;

KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;

KW cholesterol ester storage; systemic lupus erythematosus; infection;

KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;

KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;

KW bone damage; cartilage damage; antiinflammatory disease; coagulation;

KW thrombosis; contraceptive.

XX Homo sapiens.

OS WO200058473-A2.

XX PN 05-OCT-2000.

XX PD 31-MAR-2000; 2000WO-US08621.

XX PF 31-MAR-1999; 99US-0127607.

XX PR 02-APR-1999; 99US-0127636.

XX PR 05-APR-1999; 99US-0127728.

XX PR 30-MAR-2000; 2000US-0540763.

XX PA (CURA-) CURAGEN CORP.

XX PI Shinkets RA, Leach M;

XX PI WPI: 2000-602362/57.

XX DR N-PSDB; AAC75154.

XX Novel nucleic acids and peptides derived from open reading frame X,

PT useful for treating e.g. cancers, proliferative disorders,

PT neurodegenerative disorders and cardiovascular disease.

XX Claim 11; Page 1203-1206; 5507pp; English.

XX AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,

CC which represent the human OREFX open reading frames 1 to 3161. The OREFX

CC sequences have activities such as: cytostatic; hepatotropic; vulnary;

CC antipsoriatic; antiparkinsonian; nootropic; neuroprotective;

CC osteopathic; anticonvulsant; antiarthritic; immunosuppressant;

CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;

CC antidiabetic; hypotensive; dermatological; immunosuppressive;

CC antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;

CC antithyroid; and antianaemic. The sequences can be used for determining

CC the presence of or predisposition to, or preventing or treating

CC pathological conditions associated with an OREFX-associated disorder. The

CC nucleic acids can be used to express OREFX proteins in gene therapy

CC vectors. The proteins and nucleic acids may be used to treat cancers,

CC proliferative disorders, neurodegenerative disorders, osteoarthritis,

CC graft vs host disease, cardiovascular disease, diabetes mellitus,

CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus

CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,

CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,

CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,

CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance

CC coagulation; to inhibit thrombosis; and as a contraceptive.

XX

SQ Sequence 1532 AA;

Query Match 4.8%; Score 135; DB 21; Length 1532;

Best Local Similarity 21.1%; Pred. No. 0.095;

Matches 127; Conservative 61; Mismatches 237; Indels 178; Gaps 28;

Qy 22 LPSQTDGAA--RGCTGHLISSTGALGSRSLFSPLRNSMADSVDSRDIPGLPTN----- 72

Db 510 lpstsgaatrlvtgn--pstgaagt-----iprvpskvsaiqe---pgeptysshsttl 560

Qy 73 PSRLAA-----TSETCLGGEFVLHDKGPLDI---LNTQIGPSAF-----RVEVQ 115

Db 561 pkttgagaqtqwtqetgtg--eallisspsysvtgmiktatsspspmlrhtsqgitta 618

Qy 116 ADGTHAAGEKNGLEVSVTLSPQEWSSL-----QSIDTEGKNRFVFT 157

Db 619 pstnhslih-----stscpsqespaavgrhtqapqtqesqgtrsvspmdtkvtv 671

Qy 158 GGRG--GSGH-----PMVTVASDIAEARTRILAKLDPDNHGGKQKDV-----DT 200

Db 672 pgsftasghspseivpqdaptisaattfapapt-----gdghttqaptalqatpss 724

Qy 201 RSVGVGSASGIDGVSFTHSTNTSSVRSDDP-----KFWVSVGAIAAGLAGLAATGIAQA 256

Db 725 hdatlpgsggts-----lsktgaltiansvstpggpegqwisasa----- 765

Qy 257 LALTPPEPDDPTTDPDQAA-----NAAESATKDOLTQFAKNPENQKVNIDANGNAIPSG 311

Db 766 -----stspdtlaaamthtqaeasggtqts--epassgsrttsagtatpss 812

Qy 312 ELXDDIVEQIAQAKAGEVARQAQVESNAQAQRYEDQHARROEELQSSGICYGLSSA 371

Db 813 sgasgtipsgsegistsgettr---fssnps-----rdshttgattellisasahg--a 861

Qy 372 LIVAGGICAGVTTALHRRNQPAEOTTTTTTTHVYQQQTGGIPQHKVAL--MPQRRRFS 429

Db 862 ipvatgmassivpgtftlseagrptqgsptpsaspqetaalrmaqtrts 921

Qy 430 RDSQGSVASTHWSDSSEVVNPYAEEVGGARNSLSAHQPEEHYDEVAADPGYSVIONFS 489

Db 922 rgsdtislas-qatdtfstv--pptpsitssgltspqtqth----- 960

Qy 490 GSGPVTGRLIGTPGGQIQSTYALLANSGLRLGMGLTSGGET-----AYSSVNA 540

Db 961 -----tlpsgsgkfttalisnatplpvtvassastghthplhvtassvstghat 1012

Qy 541 PTP 543

Db 1013 plp 1015

RESULT 12

AAY27230

ID AAY27230 standard; Protein; 1978 AA.

XX AAY27230;

XX

DT	24-SEP-1999	(first entry)	
XX	Amino acid sequence of N. meningitidis protein ORF114-1.		
XX	Neisseria meningitidis protein; pharmaceutical; vaccine; diagnosis;		
XX	bacterial infection; treatment.		
KW	Neisseria meningitidis.		
KW	Neisseria meningitidis.		
OS	Neisseria meningitidis.		
XX	WO9936544-A2.		
PN	22-JUL-1999.		
PD	14-JAN-1999; 99WO-IB00103.		
XX	09-OCT-1998; 98GB-0022143.		
PR	14-JAN-1998; 98GB-0000760.		
PR	01-SEP-1998; 98GB-0019015.		
XX	(CHIR-) CHIRON SPA.		
XX	Grandi G, Masignani V, Pizza M, Rappuoli R, Scarlato V;		
PI	WPI; 1999-444400/37.		
XX	N-PSDB; AAX99152.		
DR	New protein and its nucleotide sequence, useful in vaccines or		
PT	diagnostic compositions for treating and/or preventing Neisseria		
PT	meningitidis infections		
XX	Claim 4; Page 88; 123pp; English.		
PS	The invention provides proteins (AAY27201-245) from Neisseria		
XX	meningitidis (strains A and B) and nucleic acid sequences (AAX99123-167)		
CC	encoding the proteins. Compositions comprising the protein, nucleic acid		
CC	or antibody specific to the protein are useful as pharmaceuticals, e.g. a		
CC	vaccine composition or a diagnostic composition. The composition is also		
CC	useful for treating or preventing an infection due to Neisseria		
CC	bacteria, especially Neisseria meningitidis.		
XX	Sequence 1978 AA;		
XX	Query Match 4.8%; Score 133.5; DB 20; Length 1978;		
XX	Best Local Similarity 22.1%; Pred. No. 0.18;		
XX	Matches 128; Conservative 69; Mismatches 230; Indels 151; Gaps 30;		
QY	57	MADSVDSRDPLGLPT-----NPSRL-----AATSTCLLGGFEVLHD 94	
Db	1072	lidgitdqeigkptykshydkaalnkprritgrtgvsihaaalddariilgaseikap 1131	
QY	95	KGPLDJI-----LNTQ-----IGPSAER-----VEVQADGT 119	
Db	1132	sgsidikahsdivleagqndaytflkctkgsgkirkckftstdhllmpapvelatang 1191	
QY	120	HAA1GEC-----NGLEVSVTLSQPSWSSLSQIDTEGKNRFVFTGGGGHPMTVA 171	
Db	1192	tlqaggnieantfrfnagpkvtivage-elqlaaegihk-----hel--- 1234	
QY	172	SDIAEATRILAKLDPDNHGRQPKD--VDRFSVGVGSA--SGIDDGVVSETHTTTNS 227	
Db	1235	-dvqksrfrigikvgksnykneinetklprvrvvaqtaatsrgwdt-vlegteftktlag 1292	
QY	228	VRSDPKFVSVGATAAGLAGLAAGTAAQALALTEPDDPTTDPQAAANAESATKDOLT 287	
Db	1293	a-----diqagvekaradakillglnvrigseeKietnsvwqkq---agrgsietlk 1345	
QY	288	QEAFKPNENKVNIDANGN----AIPSGELXDDIVEQIAQAKEAGEVARQQAVERNA--- 341	
Db	1346	lpsfesptppk--ltapgggiyvidipkgnlktei-eklakpeya--ylkqlqvaknvnw 1400	
QY	342	QAQOQYEDQHARRQEEIQLSSG-----TGYGLSSALIIVAGGTGA-GVTTALHRR 389	

[illegible]

Query Match 4.7%; Score 132.5; DB 18; Length 1026;
Best Local Similarity 21.8%; Pred. No. 0.086;

Matches 123; Conservative 64; Mismatches 222; Indels 155; Gaps 27;	
QY 26	TDGAARGGTGHLISSTGALGSRSLFPLRNSMADSVDSRDIPGLPTNPSELAATSETCL 85
DB 230	tdnaagvnlftaypsgvgstl-----slttgtdt--ltgtanndtfvagevagaat 280
QY 86	LGGEVLHDKGPDILN-----TOIGPSAFRVEVQADGTHAAIGEKNGLEVSVT 134
DB 281	ltvgdtlsggagtdvlnwqaaavaltptgtvtslmetmnvtsgaaaitntssgvtglta 340
QY 135	LSPQEWSSLSQSIDT-EGKRFVFTGGRGS-----GHPMVTVASDIAEARTILAKLDPD 188
DB 341	lntntsgaaqvttagaggnltattaaqaannvavdgranvtvas-----tgvt----- 389
QY 189	NHGRQPKVDTRSVGVGSAGIDDGVVSETHSTNTSSVRSRDPKFWVSGAIAAGLAGL 248
DB 390	-----gtttvgansaa---sgtvs---svansst-----tttgaia--vtgg 424
QY 249	AATGIAQALALTPEDDPTTTDPQAAANAESATKDQTOEAFKNPENQKVNIDANGNAI 308
DB 425	tavtvaqtag---navntltlqadvvtgnssttattvtqta-----aaatagat 470
QY 309	PSGELXD--DIVEQIAQAKEAGEVAR-----QAVESNAQAQRYEDQHARRQEELQ 359
DB 471	vagrvgavtldsaasattagkiatvtlgsfgaatidssalttnl-----t 604
QY 476	VAADPGYSVIONFS---GSGPVTGRIGTPOGQIQS-----TYALLANSQ-----GL 519
DB 605	aaaltgltvtnsvgatlgaelatg-lvftggagrdsillgattkaivmgagddtvtvssa 663
QY 520	RLGMGLTSGG---ETAVSSVNA 540
DB 664	tlgagsgvnggdgtvlnvngs 687
RESULT 14	
AA44757	
ID	AA44757 standard; Protein; 1026 AA.
XX	
AC	AA44757;
XX	
DT	04-MAY-2000 (first entry)
XX	
DE	Caulobacter crescentus surface layer protein.
XX	
KW	Surface layer protein; S-layer secretion signal; antibiotic; vaccine;
KW	recombinant fusion protein cleavage; enzyme; protein polymer;
KW	antibacterial enzyme; foodstuff.
XX	
OS	Caulobacter crescentus.
XX	
PH	Key Location/Qualifiers
FT	Cleavage-site 692..693
FT	/note= "Asp-Pro dipeptide present in S-layer secretion
FT	signal sequence. It is a site where a fusion
FT	protein comprising a target protein and the secretion
FT	signal is cleaved"
XX	
PN	WO200004170-A1.
XX	
PD	27-JAN-2000.
XX	
PF	14-JUL-1999; 99WO-CA00637.
XX	
PR	14-JUL-1998; 98CA-2237704.

(UYBR-) UNIV BRITISH COLUMBIA.	
XX	Smit J;
XX	
PI	WPI; 2000-182434/16.
XX	N-PSDB; AA250079.
DR	
XX	Cleavage of Caulobacter produced recombinant fusion proteins useful for
PT	producing vaccine peptides
XX	
PS	Example 1; Pages 21-23; 33pp; English.
XX	
CC	The patent discloses a method for cleaving a recombinant fusion protein
CC	which is produced by Caulobacter and consists of Caulobacter surface
CC	layer (S-layer) protein (containing the C-terminal secretion signal) and
CC	a target protein heterologous to Caulobacter. The cleavage of target
CC	protein from the S-layer protein is carried out under mild acid
CC	conditions so that cleavage occurs at aspartate-
CC	proline dipeptide site without solubilising the protein.
CC	An cleavage is accomplished while the fusion protein is in
CC	the insoluble aggregate form which facilitates purification of the
CC	protein. The method is useful for producing pure proteins including
CC	recombinant human and animal therapeutic antibiotic and vaccine peptides,
CC	enzymes, protein polymers, and antibacterial enzymes for foodstuffs.
CC	The present sequence is a S-layer protein from C. crescentus.
CC	The S-layer secretion signal, corresponding to the C-terminal portion of
CC	the protein from amino acid 690 onwards, is fused with a target sequence
CC	for construction of a recombinant fusion construct which is expressed in
CC	Caulobacter.
XX	
SQ	Sequence 1026 AA;
Query Match 4.7%; Score 132.5; DB 21; Length 1026;	
Best Local Similarity 21.8%; Pred. No. 0.086;	
Matches 123; Conservative 64; Mismatches 222; Indels 155; Gaps 27;	
QY 26	TDGAARGGTGHLISSTGALGSRSLFPLRNSMADSVDSRDIPGLPTNPSELAATSETCL 85
DB 230	tdnaagvnlftaypsgvgstl-----slttgtdt--ltgtanndtfvagevagaat 280
QY 86	LGGEVLHDKGPDILN-----TOIGPSAFRVEVQADGTHAAIGEKNGLEVSVT 134
DB 281	ltvgdtlsggagtdvlnwqaaavaltptgtvtslmetmnvtsgaaaitntssgvtglta 340
QY 135	LSPQEWSSLSQSIDT-EGKRFVFTGGRGS-----GHPMVTVASDIAEARTILAKLDPD 188
DB 341	lntntsgaaqvttagaggnltattaaqaannvavdgranvtvas-----tgvt----- 389
QY 189	NHGRQPKVDTRSVGVGSAGIDDGVVSETHSTNTSSVRSRDPKFWVSGAIAAGLAGL 248
DB 390	-----gtttvgansaa---sgtvs---svansst-----tttgaia--vtgg 424
QY 249	AATGIAQALALTPEDDPTTTDPQAAANAESATKDQTOEAFKNPENQKVNIDANGNAI 308
DB 425	tavtvaqtag---navntltlqadvvtgnssttattvtqta-----aaatagat 470
QY 309	PSGELXD--DIVEQIAQAKEAGEVAR-----QAVESNAQAQRYEDQHARRQEELQ 359
DB 471	vagrvgavtldsaasattagkiatvtlgsfgaatidssalttnl-----t 518
QY 360	LSSGIGYGLSSALIVAGGIGAGVTTALHRRNQAPAEQTTHVTVVOQQTGGIPOHKVAL 419
DB 519	--sgtgsj-----gigralta-----tptantltnvnglt--ttgaidtseaa- 560
QY 420	MPQERRRFSRRDSQGSVASTHWSDSSEVVNRYAEV-----GGARNSLSAHOPEHIYDE 475
DB 561	-----addgfttiniagstaasiaslaadattinisdarvtitsh-----t 604
QY 476	VAADPGYSVIONFS---GSGPVTGRIGTPOGQIQS-----TYALLANSQ-----GL 519
DB 605	aaaltgltvtnsvgatlgaelatg-lvftggagrdsillgattkaivmgagddtvtvssa 663

QY 520 RLGMGLTSGG---ETAYSSVNA 540
Db 664 tlgagsgvnggdtdvlvanvngs 687

RESULT 15
AA77177
ID AAY77177 standard; Protein; 4630 AA.

AC AAY77177;

DT 05-JUN-2000 (first entry)

DE S. venezuelae vep ORF 1 amino acid sequence #1.

XX Desosamine biosynthesis; macrolide; polyketide; methymycin; pikromycin;
KW neomethymycin; narbomycin; polyhydroxyalkanoate monomer synthase;
KW biopolymer; antibiotic; chemotherapeutic; immunosuppressant; asthma,
KW chronic obstructive pulmonary disease; respiratory inflammation;
KW hypercholesterolaemia; crop protection agent.

XX Streptomyces venezuelae ATCC15439.

OS WO200000620-A2.

PN 06-JAN-2000.

XX 25-JUN-1999; 99WO-US14398.

XX 26-JUN-1998; 98US-0105537.

XX (MINU) UNIV MINNESOTA.

PA Sherman DH, Liu H, Xue Y, Zhao L;

PI WPI; 2000-160679/14.

XX N-PSDB; AAZ87283.

DR Desosamine and macrolide biosynthetic gene clusters, useful for, e.g.

PT synthesis of methymycin and pikromycin .

PS Example 3; Figure 23; 438pp; English.

CC The invention relates to an isolated and purified nucleic acid segment
CC comprising a desosamine biosynthetic gene cluster, a fragment or its
CC biologically active variant, where the nucleic acid sequence is not
CC derived from the eryc gene cluster of Saccharopolyspora erythraea or
CC Streptomyces antibioticus. The invention also relates to a macrolide
CC biosynthetic gene cluster, or fragments thereof. The macrolide
CC biosynthetic gene cluster encodes proteins which synthesize methymycin,
CC pikromycin, neomethymycin, narbomycin or a combination of these
CC compounds. Recombinant or augmented cells comprising the desosamine
CC and/or macrolide biosynthetic gene clusters are useful for the
CC production of biologically active macrolides. The macrolide biosynthetic
CC proteins are useful for synthesis of methymycin, pikromycin,
CC neomethymycin and narbomycin. The alternative termination of polyketide
CC synthesis may be useful to prepare novel antibiotics and
CC polyhydroxyalkanoate (PHA) monomers. The compounds produced by the
CC recombinant host cells are useful as biopolymers, e.g., in packaging or
CC biomedical applications, to engineer PHA monomer synthases or to prepare
CC biologically active agents, such as chemotherapeutics,
CC immunosuppressants, agents to treat asthma, chronic obstructive pulmonary
CC disease as well as other diseases involving respiratory inflammation,
CC cholesterol-lowering agents or macrolide-based antibiotics which are
CC active against a variety of organisms, e.g., bacteria, including
CC multi-drug resistant pneumococci and other respiratory pathogens, as well
CC as viral parasitic pathogens, or as crop protection agents (e.g.,
CC fungicides or insecticides) via expression of polyketides in plants.
CC Sequences AAY77181-Y77189 and AAY77199 represent Streptomyces venezuelae
CC ATCC 15439 protein sequences. These are encoded by a DNA sequence
CC designated vep ORF 1, which actually contains 3 open reading frames. The
CC vep ORF 1 protein is defined in the specification as a PHA monomer

CC synthase.
XX Sequence 4630 AA;

Query Match 4.7%; Score 132.5; DB 21; Length 4630;
Best Local Similarity 22.2%; Pred. No. 0.68;
Matches 125; Conservative 48; Mismatches 204; Indels 187; Gaps 27;

QY 23 PSQTCGARGGTGHLISST--GALGSRSLFSPLRNSMADSV--SRDIPGLPTNPSRLA 77
Db 1156 prmhgdg-aegveghlittgtpsvmagrayqigtgtpavtdtacssslvalhavlrsir 1214
QY 78 AATSETCLLGGFEVLHDKGPLDILNTOIGPSA----FRVEVQADGTHAAIGKNGKLEVS 133
Db 1215 qgesslalatagvtmstpgmfefsrgrglaaadgrskafsdadgtswaegv--gllvve 1272
QY 134 TLSPOEWSLSQIDTEGKNRFTVGTGGRGSGHPMTVVASDIA-----EAR 178
Db 1273 rlsdae-----rnhpvlavirgsavndqgasngltapngpsq 1310
QY 179 TRILAKLDPDNHGRQPKDVD--TRSVGVSASGIDDVVSETHSTN-----S 226
Db 1311 qrvirgalad--agltpadvdaveahgtgrlg--dpieaeailgtgydrgegaplqig 1366
QY 227 SVRSDPKFWVSVCATAGLAGLAATGIAQALALTPEP---DDPTTDPQOANAASATK 283
Db 1367 silksn-----ighaqaagvgglikmvlamrhgvlprtlhvdrrpttr--vdweaggvellte 1421
QY 284 DOLTQEAFFKPNQKVNIDANGNAIPSGELXDDIVEQIAQAKEAGEVARQQAQAVESNAQA 343
Db 1422 er-----ewpetgr-----prraaisfgisgtnahivveqapeage----- 1458
QY 344 QORYEDQHARQOEELQLSSGIGYGLSSALIVAGGICAGVTTALHRRNQPAE--QFTTTT 401
Db 1459 -----aavttapeageageaadtattt 1482
QY 402 HTVVQOOTGGIPQ-----HKVALMPO--ERRFSDRRDSQGSVASTHWSDSSE 448
Db 1483 paav-----gvpeprapvvvsardaaalraqavrlrtfldgr----- 1520
QY 449 VVNYAEVGGARNLSLSAHOPEEH-----IYDEVAADPGYSVIQNFSGSG--PVTGRLIG 500
Db 1521 ---pdvtvadlgrslaartafehkaalttatraddella--gldal-----grgeqatglvlg 1571
QY 501 TFCQGIQSTYALLANSGLRLGNG 524
Db 1572 eparagrtaf-lftgggaqrvamg 1594

Search completed: September 27, 2001, 14:22:24
Job time: 120 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 27, 2001, 14:20:24 ; Search time 21.2 seconds
(without alignments)
533.212 Million cell updates/sec

Title: US-09-189-415A-2
Perfect score: 2800
Sequence: 1 MPICNLGNVNGNHLPPAP.....GETAVSSVNAAPTPGPVRFV 549

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 197339 seqs, 20590346 residues

Total number of hits satisfying chosen parameters: 197339

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents_AA.*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep.*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	141	5.0	580	2	US-08-672-564-1
2	138.5	4.9	2353	4	US-09-377-155-33
3	138.5	4.9	2353	4	US-08-913-942-4
4	134.5	4.8	1912	1	US-08-409-995-4
5	134.5	4.8	1912	3	US-08-685-467-4
6	133	4.8	2703	1	US-08-185-432-19
7	132.5	4.7	1026	1	US-08-194-230-7
8	132.5	4.7	1026	2	US-08-614-377A-7
9	132.5	4.7	1026	4	US-09-142-648B-7
10	128	4.6	1507	6	5268270-2
11	122.5	4.4	507	4	US-08-860-635A-19
12	118.5	4.2	1600	2	US-08-617-697-10
13	118	4.2	2441	1	US-08-194-488-2
14	117.5	4.2	1536	1	US-08-038-682-2
15	117.5	4.2	1536	1	US-08-302-832-2
16	117.5	4.2	1536	2	US-08-530-198-2
17	117.5	4.2	1536	2	US-08-469-880-2
18	117.5	4.2	1536	2	US-08-728-470-2
19	117.5	4.2	1536	2	US-08-617-697-2
20	117.5	4.2	1536	4	US-08-719-641-2
21	117.5	4.2	3816	4	US-09-428-517-3
22	117	4.2	675	1	US-08-386-495-10
23	117	4.2	675	5	PCT-US96-02331-10
24	117	4.2	2441	3	US-08-961-739-2
25	116	4.1	417	2	US-08-672-564-9
26	116	4.1	2890	4	US-09-413-814-67
27	116	4.1	3798	3	US-09-335-409-6

28	114.5	4.1	1186	2	US-08-861-464-8	Sequence 8, Appli
29	114.5	4.1	1186	2	US-08-396-001-8	Sequence 8, Appli
30	114.5	4.1	1186	4	US-09-323-433A-8	Sequence 8, Appli
31	114.5	4.1	1529	2	US-08-728-470-10	Sequence 10, Appli
32	114.5	4.1	1529	4	US-08-719-641-10	Sequence 10, Appli
33	114.5	4.1	5588	4	US-09-036-987A-6	Sequence 6, Appli
34	113	4.0	1004	4	US-08-916-352-2	Sequence 2, Appli
35	113	4.0	1043	3	US-08-928-361B-30	Sequence 30, Appli
36	113	4.0	1721	3	US-08-700-651-5	Sequence 5, Appli
37	113	4.0	1721	3	US-08-928-361B-6	Sequence 6, Appli
38	111.5	4.0	619	4	US-09-066-046-2	Sequence 2, Appli
39	111.5	4.0	1794	6	5183745-6	Patent No. 5183745
40	110.5	3.9	1848	4	US-08-296-791-6	Sequence 6, Appli
41	110.5	3.9	1848	5	PCT-US95-10661A-6	Sequence 6, Appli
42	110.5	3.9	2516	3	US-08-374-077C-2	Sequence 2, Appli
43	110.5	3.9	2516	4	US-08-895-590-2	Sequence 2, Appli
44	110	3.9	4551	3	US-09-320-878-1	Sequence 1, Appli
45	109.5	3.9	475	1	US-08-350-741-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1
US-08-672-564-1
; Sequence 1, Application US/08672564
; Patent No. 5824503
; GENERAL INFORMATION:
; APPLICANT: KURUME, Yoko
; APPLICANT: IZU, Hiroyuki
; APPLICANT: IZUMI, Yoshiya
; APPLICANT: SANO, Mutsumi
; APPLICANT: KATO, Ikunoshin
; APPLICANT: ITO, Makoto
; TITLE OF INVENTION: GENE ENCODING ENDOGLYCOCERAMIDASE ACTIVATOR
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Birch, Stewart, Kolasch & Birch, LLP
; STREET: P.O. Box 747
; CITY: Falls Church
; STATE: Virginia
; COUNTRY: USA
; ZIP: 22040-0747
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/672.564
; FILING DATE: 28 JUNE 1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: WEINER, Marc S.
; REGISTRATION NUMBER: 32,181
; REFERENCE/DOCKET NUMBER: 1422-0263P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 205-8000
; TELEFAX: (703) 205-8050
; TELEX: 248345
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 580
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-672-564-1

Query Match 5.0%; Score 141; DB 2; Length 580;
Best Local Similarity 23.9%; Pred. No. 0.00043;
Matches 84; Conservative 45; Mismatches 126; Indels 96; Gaps 16;

QY 4 GNLGNVNGHLLPP-----APLPSPQTDGAARGGTGHLISSTGALGRSLFSP 53
Db 215 GQGNDFANFLTFPRASAPVIGTVWAPTCQSPDTRA-----GFLNAGAPLATIOI---L 267
QY 54 RNSMADSVDSRDIPGLPTNPSR--LAAATSETCLLGGFEVLHD-----KGPLDLINTQIGP 107
Db 268 RQAVA-TVSYLDGPSAVTNGGEFTNLATVVPTPDSGVQVFTRGEDVGAPVDLVN---GK 323
QY 108 SAFRVEVQADCTHAAIGKNGLEV-----SVTLSPQEWSSLSQSDITBEGKKNRFVFTGG 159
Db 324 ASLTQSLDTDGDYAEAKFLGAEFFNPSSAAKTVTVTSDIQTITSVT-----371
QY 160 RGGSGHPMTVASDIAEARTILAKLDPMHGRQKQKDVTRSGVGSAGIDGCVVSET 219
Db 372 --GPDH-----DAYRDQPNLITAKPEPGVSGTVAFEDV--GTPVGTADVMDGGAVALP 421
QY 220 HTSTNSVSRSDPKFWYSVGAIAAGLAGLAATGTAQALAL-----TPEPDPTTDDP 271
Db 422 HTFTTNGTHR-----VIARYSG--AEGISPSVSLQYPVSVTEAPAADVATTIV 468
QY 272 DOANAASATKDOLTOEAFKNPENQK-----VNIDANGNA 307
Db 469 DPIASTAGKS---PVTTLARLDPADARGTVQFKGLGDLVLLGGPVRVDANGVA 516
RESULT 2
US-09-377-155-33
; Sequence 33, Application US/09377155
; Patent No. 6197312
; GENERAL INFORMATION:
; APPLICANT: PEAK, Ian Richard Anselm
; APPLICANT: JENNINGS, Michael Paul
; APPLICANT: MOXON, E. Richard
; TITLE OF INVENTION: NOVEL SURFACE ANTIGEN
; FILE REFERENCE: 065064/0128
; CURRENT APPLICATION NUMBER: US/09/377,155
; CURRENT FILING DATE: 1999-08-19
; PRIOR APPLICATION NUMBER: PCT/AU98/01031
; PRIOR FILING DATE: 1998-12-14
; PRIOR APPLICATION NUMBER: GB 9726398.2
; PRIOR FILING DATE: 1997-12-12
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 33
; LENGTH: 2353
; TYPE: PRT
; ORGANISM: Haemophilus influenzae
US-09-377-155-33
Query Match 4.9%; Score 138.5; DB 4; Length 2353;
Best Local Similarity 22.0%; Pred. No. 0.0064;
Matches 124; Conservative 62; Mismatches 207; Indels 171; Gaps 26;
QY 6 LGNNVNGHLLPAPPPLPSQTDGAARGGTGHLISSTGALGRSLFSPRNSMADSVDSRD 65
Db 1163 LGGTANG-----RNDGTGVINKDG-----LTLITLANGAAAGTDSN 1198
QY 66 IGPLTNPRLAANTSETCLLGGFEVLHDKGPL-----DILNTQIGPSAFRVEVQADGTHA 121
Db 1199 -----GNTISVTKDGISAGNKETITNYSALKYTKQTNT-----ADTQKEPHA 1243
QY 122 AIGKNGLEVSVTLSPQEWSSLSQSDITBEGKKNRFVFTGGRGSGHPMTVASDIAEAR--- 178
Db 1244 AV--KNAVEFV-----GKNKATVSATKDDNNGKHTVTI--DVAEAKVGD 1284
QY 179 -----TRILAKLDPMHGRQKQKDVTRSGVGSAGIDGCVVSETHST--TNSSVR--- 229
Db 1285 GLEKDTDGKIKLVNDTGDGNLLITVDA--TKGASVAKGEFNNAVTTTATTAGTNANERGV 1343
QY 230 -----SDPKFWYSVGAIAAGLAGLAATGTAQALALTPEDDPTTDDPQOANA 277
Db 1344 VVKGSGNATATETDKKKKATVGDVAKAIND--AATFV-----KVENDDSATIDDSPTDDG 1396

QY 278 AESATK--DQLTOEAFKNPENQKNVID-----ANGNAIPSGELXDDIVFQIAQOAKE 327
Db 1397 ANDALKAGDTLLKAGK---NLKVRDGNKNIITFALANDLSVKSATVSOKL-----S 1444
QY 328 AGEVARQQAQAVESNAQAQRYVEDQHARRQBELOLSSGIGYGLSSALIVAGGIGAGVTTALH 387
Db 1445 LGTNGKNVNITSDTKGLNFAKSKYKGDGDDANIHL--NGIASTLTDTL-----LNSGATTNL- 1497
QY 388 RRNPQAEQTTTTTTHTVVQOQTGGIPQHKVALMPQERRRFRDRRDSQGSVASTHWSDDSS 447
Db 1498 -----CG-----NGITDNEKKRAASVKD-----1515
QY 448 EYVNPYAEVGGARNLSAHOPEHHIYDEVAADPGYSVIONFSGSPVTGRLICTPGQGIQ 507
Db 1516 -VLNAGWNVRGVKPA--SANNQVENI--DFVAT---YDTVDFVSGDKDTSVTVESKNGKR 1569
QY 508 STYALLANSGLRLGMLTSGGE 531
Db 1570 TEVKIGAKTSVIAKHNGKLTGKE 1593
RESULT 3
US-08-913-942-4
; Sequence 4, Application US/08913942
; Patent No. 6200578
; GENERAL INFORMATION:
; APPLICANT: St. Geme, Joseph
; APPLICANT: Barenkamp, Stephen J.
; TITLE OF INVENTION: HAEMOPHILUS ADHESION PROTEINS
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr Hobbach Test Albritton & Herbert LLP
; STREET: Four Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: United States
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/913,942
; FILING DATE: 29-DEC-1997
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/409,995
; FILING DATE: 24-MAR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US96/4031
; FILING DATE: 22-MAR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Vance, Dolly A.
; REGISTRATION NUMBER: 39,054
; REFERENCE/DOCKET NUMBER: A-61053-1/RET/RMS/DAV
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2353 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-913-942-4
Query Match 4.9%; Score 138.5; DB 4; Length 2353;
Best Local Similarity 22.0%; Pred. No. 0.0064;
Matches 124; Conservative 62; Mismatches 207; Indels 171; Gaps 26;

QY	6	LGNNVNGNHLIPPAPPLPSQTDAAAGGTGHLISGTGALGSRSLFSPLRNSMADSVDSRD	65
Db	1163	LGGTANG-----RNDTGVINKDG-----LTITLANGAAAGTDA	1198
QY	66	IPGLPTNPSRLAAATSETCLGGFEVLHDHKGPL-----DILNTQIGPSAFRVEVQADGTHA	121
Db	1199	-----GNTISVTKDGISAGNKEITNVKSALTKYKDTQNT-----ADEQDKEFHA	1243
QY	122	AICEKNGLEVSVTLSPOEWSLSQSIDTEGKNRFVFTGGRGSGGHPMVTVASDIAEAR--	178
Db	1244	AV--KNAEVEFV-----GKNCAVVSAKTDNNGKHTVI--DYAEAKVGD	1284
QY	179	-----TRILAKLDPNHGGRRQPKVDTRSVGVSAGIDGVMVSETHST--TNSSVR---	229
Db	1285	GLEKDTDGGIKLKVDNTDGNLLTVDA--TKGASVAKGEFNAVTTDATTAQGTNANERGV	1343
QY	230	-----SDPKFWVSUCAIAGLAGLAANTCIAQALALTEPDDPTTTPDQAANA	277
Db	1344	VKGSNGATATETDKKKVATVGDVAKAIND--AATFV-----KYENDDSATIDDSPDDG	1396
QY	278	AEASATK--DQLTQEAFFKNPENOKVID-----ANGNAIPSGELXDDIVEQIAQAQAKE	327
Db	1397	ANDALAKAGDTLLKAGK---NLKVRDGENITFALANDLSVKSVATVSDKL-----S	1444
QY	328	AGEVARQQAVESNAQAQRYEDOHARROEEQLQSSGIGYGLSSALIIVAGGIGAGVTTLAH	387
Db	1445	LGTNGKNVITSDTKGLNFAKDKSGTDDANIHL--NGIASTLTIDL-----LNSGATTNL-	1497
QY	388	RRNQPAEQTTTTTHTVVQQTGPIQHKVAMPQERRRFSDDRDSOGSVASTHWSOSS	447
Db	1498	-----GG-----NGITDNEKKRAASVKD-----	1515
QY	448	EVVNPVAYEGGARNLSAHOPEEHYDEVAADPGVSVIQNFSGSGPVTRLIGTPGOGIQ	507
Db	1516	-VLNAGWNVRGVKPA--SANNQVENT--DFVAT---YDVFVSGDKDTSVTVESKDNKR	1569
QY	508	STYALLANSGLRLGMGGILTSGGE	531
Db	1570	TEVKIGAKTSVIKDHNGKLFSGKE	1593
RESULT	4		
US-08-409-995-4			
Sequence 4, Application US/08409995			
Patent No. 5646259			
GENERAL INFORMATION:			
APPLICANT: Barenkamp, Stephen I.			
APPLICANT: St. Geme III, Joseph W.			
TITLE OF INVENTION: Haemophilus Adhesion Proteins			
NUMBER OF SEQUENCES: 6			
CORRESPONDENCE ADDRESS:			
ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert			
STREET: Four Embarcadero Center, Suite 3400			
CITY: San Francisco			
STATE: CA			
COUNTRY: USA			
ZIP: 94111-4187			
COMPUTER READABLE FORM:			
MEDIUM TYPE: Floppy disk			
COMPUTER: IBM PC compatible			
OPERATING SYSTEM: PC-DOS/MS-DOS			
SOFTWARE: PatentIn Release #1.0, Version #1.30			
CURRENT APPLICATION DATA:			
APPLICATION NUMBER: US/08/409,995			
FILING DATE: 24-MAR-1995			
CLASSIFICATION:			
ATTORNEY/AGENT INFORMATION:			
NAME: Silva, Robin M.			
REGISTRATION NUMBER: 38,304			
REFERENCE/DOCKET NUMBER: A-61053/RFT			
TELECOMMUNICATION INFORMATION:			
TELEPHONE: (415) 781-1989			

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/685,467
FILING DATE: 22-JUL-1996
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/409,995
FILING DATE: 24-MAR-1995
ATTORNEY/AGENT INFORMATION:
NAME: Silva, Robin M.
REGISTRATION NUMBER: 38,304
REFERENCE/DOCKET NUMBER: A-61053-2/RFT/RMS
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
TELEX: 910 277299
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1912 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-685-467-4

Query Match 4.8%; Score 134.5; DB 3; Length 1912;
Best Local Similarity 22.0%; Pred. No. 0.01;
Matches 124; Conservative 62; Mismatches 207; Indels 171; Gaps 26;
QY 6 LGNNVNGNHLIPPPALPSQTDGARGGTGHLISGTALGSRSLFSPLRNSMADSVSRD 65
DB 1162 LGGTANG-----RNDTGVINKDG-----LTITLANGAAAGTDA 1197
QY 66 IPGLTPNPSRLAAATSETCLGGFEVLHDKGPL-----DILNTOIGPSAFRVEVQADGTHA 121
DB 1198 -----GNTISVTKDGISAGNEITNVKALKYKDTQNT-----ADETDQKEFHA 1242
QY 122 AIGKNGLEVSVTLSPQEWSSLSQSDTEGKRNRFVFTGGRGGSGHPMVTVASDIABAR--- 178
DB 1243 AV--KNAENEVEFV-----GKNGATVSAKTDNNGKHVTI--DVAAEKVGD 1283
QY 179 -----TRILAKIDPDNHHGRQPKVDYTRSVGVGSAGIDGCVWSETHST--TNSSVR--- 229
DB 1284 GLEKTDGKIKLVKVDNTDGNLLTVDA--TKGASVAKGEFNAVTTDAITAQGTNANERKV 1342
QY 230 -----SDPKFVWSVGAIAAGLAGLAATGIAQALALTPPEPDDPTTDDPQAAANA 277
DB 1343 VYKGSNGATATETDKKKVATGVDVAKAIND--AATFV-----KVENDDSATIDDSPTDDG 1395
QY 278 AESATK--DQLTOEAFKPNQKVNID-----ANGNAIPSGELXDDIVEQTAQQAQKE 327
DB 1396 ANDALKAXDITLTKAGK--NLKVRKRDGNITFALANDLSVKASATVSKL-----S 1443
QY 328 ACEVARQAQAVESNAQAQRYEDQHARRQEEQLQSSGIGYGLSSALIVAGGICAGVTTALH 387
DB 1444 LGTNGKNVNTSDTKGLNFAKDSKTDGDANHL--NGIASTLTDIL-----LNSGATTNL- 1496
QY 388 RRNPQAEQTTTTTHTVVOQQGGIPQHKVALMPQERRFRDRRSQCSVASVTHWSDS 447
DB 1497 -----GG-----NGITDNEKKXAA-----SVK 1513
QY 448 EVNYPYAEVGGARNLSAHOPEHIYDEVAADPGYSVIONFSGSGPVTGRLLGTGPGCQIQ 507
DB 1514 DVLNAGWNVRGVKPA--SANNQVNTI--DFVAT---YDVTDFVSGDKDTTSVTESKDNKR 1568
QY 508 STYALLANSGLRLGMGGITSGGE 531
DB 1569 TEVKGAKTSVTKDHNGKLFKGE 1592

RESULT 6
US-08-185-432-19
Sequence 19, Application US/08185432
Patent No. 5750652
GENERAL INFORMATION:
APPLICANT: Artavanis-Tsakonas, Spyridon
APPLICANT: Busseau, Isabelle
APPLICANT: Diederich, Robert J.
APPLICANT: Xu, Tian
APPLICANT: Matsuno, Kenji
TITLE OF INVENTION: DELTEX PROTEINS, NUCLEIC ACIDS, AND
TITLE OF INVENTION: ANTIBODIES, AND RELATED METHODS AND COMPOSITIONS
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSEE: PENNIE & EDMONDS
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/185,432
FILING DATE: 21-JAN-1994
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Misrock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 7326-006
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 19:
SEQUENCE CHARACTERISTICS:
LENGTH: 2703 amino acids
TYPE: amino acid
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-185-432-19

Query Match 4.8%; Score 133; DB 1; Length 2703;
Best Local Similarity 19.8%; Pred. No. 0.024;
Matches 119; Conservative 83; Mismatches 193; Indels 206; Gaps 28;
QY 4 GNLGNNVNGNHLIPPPALPSQT--DGAAR-----GGTGHLSSTGALGSRSLFSPLRN 55
DB 2177 GNGGNNNGN---ASGKQSNQAKQAQKAKLIEGSPDNGLDATGSLRRK---ASSKK 2229
QY 56 SMADSVSRDIPGLTPNPSRLAAATSETCLGGFEVLHDKGPLDILNTOIGPSAFRVEVQ 115
DB 2230 TSAASKAANLNGL--NPGQLT-----GGVSGVPGVPP---TNSAVQAAAAAAAV 2275
QY 116 ADGTHAAIGKNGLEVSVTL--SPOEWSLSQSIDTEGKRNRFVFTGGRGGSGHPMVTVASDI 174
DB 2276 AAMSHLEGSVPVGMGNLPSPYDTSSMY----- 2306
QY 175 AEARTILAKIDPDNHHGRQPKVD-----TRSVGVGSASGIDGCVVSEHTSTNNSVR 229
DB 2307 -NAMAAPLANGNP--NTGAKOPPSYEDCICKNAQSMQSLQGNGL--DMIKLDNAYSMGSPFQ 2363
QY 230 SD-----PRFWVSVGAIAAGLAGL-----AATG----- 252
DB 2364 QELLNGOGLMNGNQQRNGVPGVLPGLCGMGLSGAGNNSREQGLSPYPYSNQSPHS 2423
QY 253 IQAALALTP-----PDDPTTDPDQAAANAAESATKDQLTQEAF----- 291

REGISTRATION NUMBER: 34053
REFERENCE/DOCKET NUMBER: 08106/002001
TELEPHONE: 617-542-5070
TELEFAX: 617-542-8906
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 1026 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-614-377A-7

Query Match 4.7%; Score 132.5; DB 2; Length 1026;
Best Local Similarity 21.8%; Pred. No. 0.0059;
Matches 123; Conservative 64; Mismatches 222; Indels 155; Gaps 27;

QY 26 TDGAARGGTGHLISSTGALGSRSLFSLRNSMADSVDRDIPGLTPNPSRLAAATSETCL 85
DB 230 TDNAAGVNLFTAYPSSGVSGSTL-----SLTGTGDT--LTGTANNDTFVAGEVAGAA 280

QY 86 LGGFVFLHDKGPLDILN-----TGIGPSAFRVEQADGTHAAIGKNGLEVSVT 134
DB 281 LTVGDTLSGGAGTDVNLVWQAAVATLPTGTISGIETMNTVTSAAITLNTSSGVTGLTA 340

QY 135 LSPQWSSLSQIDT--EGKNRFVFTGGRGS-----GHPMVTVASDIAEARTILAKLDPD 188
DB 341 LNTNTSGAAQTVTAGAGQNLTTATTAQAANNVAVDGRANVTVAS-----TGVTS----- 389

QY 189 NHGGROPKDVDRSVGSGASGIDGVVSETHSTNTSSVSRSDPKFVWVGAIAGLAGL 248
DB 390 -----GTTTVGANSAA---SGTVS---VSVANSST-----TTTGAIA--VTGG 424

QY 249 AATGIAQALALTPEDPTTDPDQAANAASATKQDTQEAFAKNPENQKYNIDANGNAI 308
DB 425 TAVTVAQTAG---NAVNTLTQADTVTGNSTTAVTTQTA-----AATAGAT 470

QY 309 PSGLXD--DIVEQIAQAQAKEAGEVAR-----QQAVERNAQAQORYEDQHARRQEELQ 359
DB 471 VAGRVNGAVTITDSAAASATAGKIATVTLGSGAATIDSSALTIVNL----- 518

QY 360 LSSGIGYGLSSALIVAGGIGACVTTALHRRNQAPAEQTTTTTHVVQOQTGGIPQHKVAL 419
DB 519 --SGTGTSL-----GIGRGALTA-----TPTANTLTNLVNGLT--TTGAIITDSEAA- 560

QY 420 MPQERRFRDRDSQGSVASTHWSDSSESVVNPYAEV---GGARNSLSAHOPEEHIYDE 475
DB 561 -----ADGFTTINIAGSTASTIASLVAADATTLNISGDARVTITSH-----T 604

QY 476 VAADPGYSVIONFS---GSGPVTGRIGTPGQGIQS-----TYALLANS-----GL 519
DB 605 AALTGITVNSVGTALGAEATG-LVFTGGAGRDSILLGATTKAIVMGAGDDTFTVSSA 663

QY 520 RLGMGLTSGG-----ETAVSSVNA 540
DB 664 TLGAGGSVNGGDTDLVAVNNGS 687

RESULT 9
US-09-142-648B-7
Sequence 7, Application US/09142648B
Patent No. 6210948
GENERAL INFORMATION:
APPLICANT: Smit, John
APPLICANT: Bingle, Wade H.
APPLICANT: No. 6210948ellini, John F.
TITLE OF INVENTION: EXPRESSION AND SECRETION OF HETEROLOGOUS POLYPEPTIDES FROM CAULO
FILE REFERENCE: 08106/002002
CURRENT APPLICATION NUMBER: US/09/142,648B
CURRENT FILING DATE: 1999-03-30
PRIOR APPLICATION NUMBER: PCT/CA97/00167

PRIOR FILING DATE: 1997-03-10
PRIOR APPLICATION NUMBER: US 07/614,377
PRIOR FILING DATE: 1996-03-12
NUMBER OF SEQ ID NOS: 12
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 7
LENGTH: 1026
TYPE: PRT
ORGANISM: Caulobacter crescentus
US-09-142-648B-7

Query Match 4.7%; Score 132.5; DB 4; Length 1026;
Best Local Similarity 21.8%; Pred. No. 0.0059;
Matches 123; Conservative 64; Mismatches 222; Indels 155; Gaps 27;

QY 26 TDGAARGGTGHLISSTGALGSRSLFSLRNSMADSVDRDIPGLTPNPSRLAAATSETCL 85
DB 230 TDNAAGVNLFTAYPSSGVSGSTL-----SLTGTGDT--LTGTANNDTFVAGEVAGAA 280

QY 86 LGGFVFLHDKGPLDILN-----TGIGPSAFRVEQADGTHAAIGKNGLEVSVT 134
DB 281 LTVGDTLSGGAGTDVNLVWQAAVATLPTGTISGIETMNTVTSAAITLNTSSGVTGLTA 340

QY 135 LSPQWSSLSQIDT--EGKNRFVFTGGRGS-----GHPMVTVASDIAEARTILAKLDPD 188
DB 341 LNTNTSGAAQTVTAGAGQNLTTATTAQAANNVAVDGRANVTVAS-----TGVTS----- 389

QY 189 NHGGROPKDVDRSVGSGASGIDGVVSETHSTNTSSVSRSDPKFVWVGAIAGLAGL 248
DB 390 -----GTTTVGANSAA---SGTVS---VSVANSST-----TTTGAIA--VTGG 424

QY 249 AATGIAQALALTPEDPTTDPDQAANAASATKQDTQEAFAKNPENQKYNIDANGNAI 308
DB 425 TAVTVAQTAG---NAVNTLTQADTVTGNSTTAVTTQTA-----AATAGAT 470

QY 309 PSGLXD--DIVEQIAQAQAKEAGEVAR-----QQAVERNAQAQORYEDQHARRQEELQ 359
DB 471 VAGRVNGAVTITDSAAASATAGKIATVTLGSGAATIDSSALTIVNL----- 518

QY 360 LSSGIGYGLSSALIVAGGIGACVTTALHRRNQAPAEQTTTTTHVVQOQTGGIPQHKVAL 419
DB 519 --SGTGTSL-----GIGRGALTA-----TPTANTLTNLVNGLT--TTGAIITDSEAA- 560

QY 420 MPQERRFRDRDSQGSVASTHWSDSSESVVNPYAEV---GGARNSLSAHOPEEHIYDE 475
DB 561 -----ADGFTTINIAGSTASTIASLVAADATTLNISGDARVTITSH-----T 604

QY 476 VAADPGYSVIONFS---GSGPVTGRIGTPGQGIQS-----TYALLANS-----GL 519
DB 605 AALTGITVNSVGTALGAEATG-LVFTGGAGRDSILLGATTKAIVMGAGDDTFTVSSA 663

QY 520 RLGMGLTSGG-----ETAVSSVNA 540
DB 664 TLGAGGSVNGGDTDLVAVNNGS 687

RESULT 10
5268270-2
Patent No. 5268270
APPLICANT: Meyer, Thomas F.; Halter, Roman; Pohlner, Johannes
TITLE OF INVENTION: PROCESS FOR PRODUCING PROTEINS USING GRAM
NEGATIVE HOST CELLS
NUMBER OF SEQUENCES: 6
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/171,872
FILING DATE: 01-JUL-1987
SEQ ID NO: 2
LENGTH: 1507
5268270-2

Query Match 4.6%; Score 128; DB 6; Length 1507;

;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/617,697
;; FILING DATE: 01-APR-1996
;; CLASSIFICATION: 424
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 08/302,832
;; FILING DATE: 05-OCT-1994
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US PCT/US93/02166
;; FILING DATE: 16-MAR-1993
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Berkstresser, Jerry W
;; REGISTRATION NUMBER: 22,651
;; REFERENCE/DOCKET NUMBER: 1038-557
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (703) 415-0810
;; TELEFAX: (703) 415-0813
;; INFORMATION FOR SEQ ID NO: 10:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 1600 amino acids
;; TYPE: amino acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; US-08-617-697-10

Query Match 4.2%; Score 118.5; DB 2; Length 1600;
Best Local Similarity 19.4%; Pred. No. 0.2; Indels 157; Gaps 26;
Matches 124; Conservative 84; Mismatches 275;
QY 1 MPIGNL--GN-----NVNGNHLPPPLPSQTDGAARGGTGHLISSTGALGSRSL 49
Db 878 LFLGNISVEGNLSLGCANANVGN-----LSTAEDSTKGEASDNLNITGTFTNGT 929
QY 50 FSLPLNSMADSVDSRDELPGLPNPRLAAATSETCLLGGFEVLHDKGDLIDLN-----T 103
Db 930 ANINIKGVVVLGINNKGGL--NITNASGTQKTIING--NITNEKGLDNKINKADAEI 985
QY 104 QIGPSAFVE---VQADGTHAA--IGEKNGLE-----VSVTLSPQEWSSLSQSI 146
Db 986 QIGGNISQKEGNLTSSDKVNTNQTITKAGVEGGRSDSSAEANLTIQKELAGDL 1045
QY 147 DREGKRRFVFTGGGGGHPMVTVASDIAEARTRIL-----AKLDPDNHGGKQPKVDVTR 201
Db 1046 NISGFNKAEIT-AKNGSDLTICNAGSNADAKKVTDFKDKSKISTDGHNVTLNSEVKT 1104
QY 202 SVGVGSASIDGQVYSEHTSTNSVSRDPKFWVSVA-----IAAGLAGLAATGIAQAL 257
Db 1105 NGSSNAGNDNSTGLTISAKDKVTNNNNVTSKHTINISAAAGNVTTKEGTTINATTSVEVT 1164
QY 258 ALTPPEDDPTTDPQAAANAESATKDQLTQE-AFKNPENQKVN1-----D 302
Db 1165 A-----QNGTIKGNITSQNVTVTATENLVTTENAVINATSGTVNISTKTDIKGGIEST 1219
QY 303 ANGNATPSGEL--XDDIVQIAQAAKEAEVARQ-----QAVESNAQAQRYEDQHARRQ 355
Db 1220 GNVNITASGNLTKVSNITQDVTVTADAGALATTAGSTISATTGNANITTKTGDKNGK-- 1277
QY 356 EELQLSSG-IGYCLLSALIVAGGIGAG---VTTALHRRNQPAEQTTTTTHVWQOQTGG 411
Db 1278 --VESSSGSVTLVATGATLAVGNISNTVTTADSKLSTVSGTSTNGTNSVTSQSOG- 1334
QY 412 IPQHKVALMPQERRRFSRRDSQGSVASTHWSDDSEVVNPNYAEVGGARNLSAHOPEH 471
Db 1335 -----DIEGTI-----SGNTVNTASTGDLTIGNSAKVEAKN 1366
QY 472 IYDEVAADPGYSVIQNFSG-----SGPV-----TGRLLGTTPGQG 505
Db 1367 GAATLTAESKLTATGCGSSITSSNGQTTITAKDSSSTAGNINAANVTLNTGTTLTTTGDSK 1426
QY 506 IOSTYALLANGSLRLG-----MGSLTSGGETAVSSVNA 540
Db 1427 INAT-----SGTLTINAKDAKLDAASGDRTVVNATNAS 1460

RESULT 13
US-08-194-468-2
; Sequence 2, Application US/08194468
; Patent No. 5750336
; GENERAL INFORMATION:
; APPLICANT: Montminy, Marc R.
; TITLE OF INVENTION: ASSAYS FOR THE IDENTIFICATION OF
; TITLE OF INVENTION: COMPOUNDS WHICH INHIBIT ACTIVATION OF CAMP AND MITOGEN
; TITLE OF INVENTION: RESPONSIVE GENES
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pretty, Schroeder, Brueggemann & Clark
; STREET: 444 South Flower Street, Suite 2000
; CITY: Los Angeles
; STATE: California
; COUNTRY: USA
; ZIP: 90071
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/194,468
; FILING DATE: 10-FEB-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Reiter, Stephen E.
; REGISTRATION NUMBER: 31,192
; REFERENCE/DOCKET NUMBER: P41 9672
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619)-546-4737
; TELEFAX: (619)-546-9392
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2441 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-194-468-2

Query Match 4.2%; Score 118; DB 1; Length 2441;
Best Local Similarity 19.0%; Pred. No. 0.43;
Matches 101; Conservative 70; Mismatches 177; Indels 184; Gaps 26;
QY 1 MPIGNLGNVNGNHLPPPLPSQTDGAARGGTGHLISSTGALGSRSLFSPLRNSMADS 60
Db 45 IPNGEL-SLLNSGNLVPDAASKHQLSELLRGGSGSSINP--GIGNVSASSPVQOGLGGQ 101
QY 61 VDSR-----DIPGL-----PTNPSRLA----- 77
Db 102 AOGQPNSTMASLGMKGKSPLNQGDSSSTPNLPKQAASSTSGPTPPASQALNPQAOKQVGLV 161
QY 78 ---AATST---CLLGGFEVLHDKGDLIDLNTQIGPSAFRVEVQA-----DGTAAIGE 125
Db 162 TSSPATSTQGTGICMNAFNQTHP---GLLNSNSGHSMLNQAQOQOQAVNNGSLGAAGR 217
QY 126 KNG-----LEVSVTLSPQ---EWSLSQSIDTECKNRFVFTGGRGSGSH 165
Db 218 GRGAGMPYPAPAMOGATSSVLAETLTQVSPOMAGHAGLNTAAGGTMKGMTGTTSPFGQ 277
QY 166 PMVTVASDIAEARTRILAKLPDNHGGKQPKVDVTRSVGVGSASIDGQVYSETH----- 220
Db 278 PF-----SQTGGQOM-----GATGVNPQLASKQSMVNSL 306
QY 221 ----TSTNNSVSRDP---KFWVSVAIAAGLAGLAATGIAQALALTPEDDPTTDPDQ 273
Db 307 PAFPTDIKNTSVTTVPNNMSQLTQTSVGIYP-----TQAIATGP-----TADPE- 348
QY 274 AANAASATKDQLTQE-----AFKNPENQKVNINDANGNAIPSGELXDDIVEQIAQ- 325

Db 349 -----KRKLIQQQLVLLHAKCQRREQANCEVRACSLPHCRMTKKNVNLNHTHQA 399
QY 326 KEAGVAR---QOAVESNAQAQRYE-----DOHARROEELQLSSGIGYGLSSALIV 374
Db 400 PKACQVAHCASSRQIISHWKNCTRHDPCVCLPLKNASDKRNQQTILGSPAS-GIQNTI-- 456
QY 375 AGGIGAGVTTALHRRN-OPABOTTTTTTHVVQQTGGIP---OHKVALMPQ 422
Db 457 -GSVGAGQONATSLNPNPIDPSSMORAYAL-----GLPYMNQDTQLQFO 502

RESULT 14
US-08-038-682-2
; Sequence 2, Application US/08038682
; Patent No. 5549897
; GENERAL INFORMATION:
; APPLICANT: BARENKAMP, STEPHEN J
; APPLICANT: ST. GEME III, JOSEPH W
; TITLE OF INVENTION: HIGH MOLECULAR WEIGHT SURFACE PROTEINS
; TITLE OF INVENTION: OF NON-TYPEABLE HAEMOPHILUS
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Mattare, Ltd
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; STREET: Bldg. 1
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/038,682
; FILING DATE: 16-MAR-1993
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: BERKSTRESSER, JERRY W
; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: 1038-293
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0813
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1536 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-038-682-2

Query Match 4.2%; Score 117.5; DB 1; Length 1536;
Best Local Similarity 19.7%; Pred. No. 0.23; 239; Indels 213; Gaps 26;
Matches 129; Conservative 73; Mismatches 239; Indels 213; Gaps 26;
QY 37 LISSTGALGSRSLFPLRNSMADSVSDRDPGLPTNPRLAAATSETCLLGGF----- 89
Db 223 VLSVNG--GSISLAGQKTTISDIIN-----PTTYSIAAPENAVNLGDIKAGGNI 273
QY 90 -----EVLHDKGPDILNTQIGPSAFRVEQADTHAAIG----- 125
Db 274 NVRAATIRNQGLSDSVSKDSGNVLSAKEGEAEIGGVISAQQAQKGLMTGDKV 333
QY 126 -----KNGLEV-----SVTLSPQE-----W 140
Db 334 TLTKGAVIDLSCKEGGTYLGDGERGKNGIQIAKKTSLKSGSTINVSKEKGRAIWW 393
QY 141 SSLQSID----TEGKNRFVFTGG-RGGSGHPMVTVASDIAEARTRILAKLPDN----- 189

Db 394 GDIALIDGNIQAQSGGDIKATGGFVETSGHDLFIKDNAIVDAKEWL-----LDFDNVSNAE 450
QY 190 HGGQPKDQVDRSVGVG--SASGIDGCVVSETHSTTNSNSVRSRDPKFW----- 235
Db 451 TAGSRNTSEDEYTGSGNSASTPKRNKEKTTLTNTLESILKKGTFVNTANQRYYVNS 510
QY 236 --VSVGATAAGLAGLAATGIAALALTPEDDPTTDPDQAANAASATKDQLT----- 287
Db 511 INLSNGSLTLWSEGRSGGV-----EINNDITGDDTRGAN-----LTIYSGW 554
QY 288 QEAFKNPENQKVNIDANGNAIPSGELXDDIVEQIAQAQAKEAGEVARQQAQVESAQAQRY 347
Db 555 VDVHKN-----ISLGAQONI-----NITAKODIAFEKGSNOVITGQ----- 591
QY 348 EDQHARQOEELQSSGIGYGLSSALIYAGGIGAGVTTALHRRNQA-----EQTTT----- 398
Db 592 -----TITSGNQKGFNNVSLNGTSGLGQFTTKRTNKYAITNKFEGLNISK 640
QY 399 TTTHTVVQQTGGIPQHK-----VALMPOERRRFSRRDSQGSVASTHWSDSSEVVN 451
Db 641 VNISMVLPKNESGYDKFKGRTYWNLTSLNVSSEGEFNLITDSRG-----SDSAGTLTQ 693
QY 452 PYAEVGGARNSLSAHQPEEH--IYDEVAADPG---YSVIQ-----NFGSGPVTGRL 498
Db 694 PYNLNGISFNKDTTFNVERNARVNFIDIKAPIGINKYSSLNVSFNGNISVSGGSDVDFTL 753
QY 499 IG-----TPCGGIQSTYALLANSGLRLGMGLTSGGETAVSSVNAAPTGPV 546
Db 754 LASSSNVQTPGVVINSKYFNVSTGSSLRFTSGSTKTFGSIEKDLTLNATGNI 807

RESULT 15
US-08-302-832-2
; Sequence 2, Application US/08302832
; Patent No. 5603938
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen J
; TITLE OF INVENTION: High Molecular Weight Surface Proteins
; TITLE OF INVENTION: of No. 5603938-Typeable Haemophilus
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Mattare, Ltd.
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; STREET: Bldg. 1
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/302,832
; FILING DATE: 16-SEP-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9205704.1
; FILING DATE: 16-MAR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US pct/us93/02166
; FILING DATE: 16-MAR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Berkstresser, Jerry W
; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: 1038-404
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0813
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:

```

; LENGTH: 1536 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-302-832-2

Query Match 4.2%; Score 117.5; DB 1; Length 1536;
Best Local Similarity 19.7%; Pred. No. 0.23;
Matches 129; Conservative 73; Mismatches 239; Indels 213; Gaps 26;

Qy 37 LISSTGALGSRSLFSPURNMADSVDSRDIPGLPTNPRLAAATSETCLLGGF----- 89
Db 223 VISVNG--GSISLLAGQITISDIIN-----PTIYSIAAPENEAVNLGDIFAKGGNI 273
Qy 90 -----EVLHDKGLDILNTOIGPSAFRVEVQADGTHAAIGE----- 125
Db 274 NVRAATIRNOGKLSADSVSKDKNIVLSAKEGEAIGGVISAQNOQAKGKLMITGDKV 333
Qy 126 -----KNGLEV-----SVTLSPQE-----W 140
Db 334 TLKTGAVIDLSGKEGGTYLGGDERGEGKNGIQIAKTSLEKSTINVSKEKGGRRAIVW 393
Qy 141 SSLQSID----TEGKNRFVFTGG--RGGSGHPMTVVASDIAEARTILAKLDPDN----- 189
Db 394 GDIALIDGNINAQSGGDIAGTGGFVETSGHDLFKDNAIVDAKEWL---LDFDNVSINAE 450
Qy 190 HGGRPQKDVTRSVGVG--SAGSIDDGVVSETHSTTNSSVRSDFKFW----- 235
Db 451 TAGRSNTSEDEXTYTGSGNSASTPKRNEKKTTLTNTLESILKKGTFVNTANQRIYVNSS 510
Qy 236 --VSVGATAAGLAATGIAQALALTPEDDPTTDPDQAAANAESATKDQLT----- 287
Db 511 INLSNGSLTLWSEGRSGGV-----EINNDITGDDTRCAN-----LTIYSSGW 554
Qy 288 QEAFKPNENKVNIDANGNAIPSGELXDDIVEQIAQAKAGEVAROQAVESNAQAQORY 347
Db 555 VDVHKN-----ISLGAQONI-----NITAKQDIAFEKGSNQVITGQG----- 591
Qy 348 EDQHARQEEQLSSGIGYGLSSALIVAGGIGAGVTTALHRRNQPA-----EQTTT--- 398
Db 592 -----TITSGNQKQFRFNNVSLNGTSGGLQFTTKRTNKYAITNKEGTLNISGK 640
Qy 399 TTHTVVOOQTGGIPOKH-----VALMPOERRRFRDRDSQGSVASTHWSDSSESVVN 451
Db 641 VNISWLPKNEGYDKFKGRTYWNLTSLNVSEGEFNLITDSRG-----SDSAGTLTQ 693
Qy 452 PYAEVGGARNLSAHQPEEH--IYDEVAADPG---YSVIQ-----NFSGSGPVTGRL 498
Db 694 PYNLNGISFNKDDTTFNVERNARVNFIDKAPIGINKYISLNYASFNGNISVSGGSDVFTL 753
Qy 499 IG-----TPQGGIQTSTYALLANSGLRLGMGLTSGGETAVSSVNAAPTGPV 546
Db 754 LASSSNVQTPGVVINSKYFNVTGSSSLRFKTSKTKTGFSTEKDLTLNATGGNI 807
```

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 27, 2001, 14:20:24 ; Search time 25.62 Seconds
(without alignments)
1632.312 Million cell updates/sec

Title: US-09-189-415a-2

Perfect score: 2800

Sequence: 1 MPICNLGNVNGNHLIPAP.....GETAVSVNAAPTGPVRFV 549

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR_68:*

1: pir1:*

2: pir2:*

3: pir3:*

4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1528.5	54.6	558	E86045	probable transloca
2	150.5	5.4	1608	A28182	hemolysin A - Serr
3	150	5.4	1005	C71513	hypothetical prote
4	149.5	5.3	2806	D85644	hypothetical prote
5	149	5.3	2232	T34434	hypothetical prote
6	146	5.2	1829	T24583	hypothetical prote
7	144.5	5.2	1296	T13936	collar protein iso
8	144	5.1	883	S04722	puif 74E protein -
9	144	5.1	1589	T42233	submaxillary mucin
10	141.5	5.1	647	T39141	hypothetical prote
11	140	5.0	2132	A55182	aggreccan precursor
12	137	4.9	1532	A26039	iga-specific metal
13	137	4.9	2468	A83412	hypothetical prote
14	136.5	4.9	770	T51024	related to C2H2 zi
15	136.5	4.9	1829	T35681	probable sensory h
16	136	4.9	674	B82117	probable flagellar
17	136	4.9	1569	A65044	hypothetical prote
18	135.5	4.8	2124	A28452	proteoglycan core
19	135	4.8	1085	S66149	gene pipsqueak pro
20	134.5	4.8	13288	T03099	mucin, submaxillar
21	133.5	4.8	1975	B81192	hemagglutinin/hemo
22	133.5	4.8	1995	B81044	hemagglutinin/hemo
23	133	4.8	3442	E82589	hemagglutinin-like
24	133	4.8	3455	B82515	hemagglutinin-like
25	132.5	4.7	812	T48016	probable zinc-fing
26	132.5	4.7	1026	T48995	paracrystalline su
27	132.5	4.7	1630	A43577	ascites sialoglyco
28	132	4.7	666	G82111	flagellar hook-ass
29	132	4.7	827	A37849	S-layer protein -

30	132	4.7	1528	2	D85912	hypothetical prote
31	132	4.7	5188	2	B85547	probable RTX famil
32	131.5	4.7	635	2	F75477	hypothetical prote
33	131.5	4.7	1804	2	H96597	hypothetical prote
34	131	4.7	2703	1	A24420	notch protein - fr
35	130.5	4.7	786	2	T16509	hypothetical prote
36	130.5	4.7	930	2	D37271	A-alpha 2 4 protei
37	130.5	4.7	1018	2	H83135	probable adhesin p
38	130.5	4.7	2422	2	T12687	ALR protein homolo
39	130	4.6	1596	2	A33106	neurogenic locus m
40	130	4.6	3190	2	T13828	CREB-binding prote
41	129	4.6	1280	2	T00365	hypothetical prote
42	129	4.6	2415	1	A39086	aggreccan precursor
43	128.5	4.6	618	2	A42020	regulatory protein
44	128.5	4.6	1536	2	A43855	high-molecular-wei
45	128	4.6	827	2	JC4900	xanthomonapepsin (

ALIGNMENTS

RESULT 1

E86045

probable translocated intimin receptor protein tir [imported] - Escherichia coli (str C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 31-Mar-2001
C:Accession: E86045

R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; May
Miller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apoda
Nature 409, 529-533, 2001

A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.

A:Reference number: A85480; MUID:21074935; PMID:11206551

A:Accession: E86045

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-558 <STO>

A:Cross-references: GB:AE005174; NID:g12518449; PIDN:AAG58825.1; GSPDB:GN00145; UWGP:

A:Experimental source: strain O157:H7, substrain EDL933

C:Genetics:

A:Gene: tir

Query Match	54.6%	Score 1528.5;	DB 2;	Length 558;
Best Local Similarity	56.2%	Pred. No. 6e-88;		
Matches 327;	Conservative 60;	Mismatches 138;	Indels 57;	Gaps 11;
QY	1	MPICNLGNVNGNHLIPAPLPSPQTDGAARGTGHLSSTGALGSRSLFSPLRNSMADS	60	
Db	1	MPICNLGNVNGNHLIPAPLPSPQTDGA--GORGOLINSTGTLGSRALFTPVNSMADS	58	
QY	61	VDSR--DIPGLTNPRLAAATSETCLLGGFEVLHDKGLDILNTQIGPSARVEVQADG	118	
Db	59	GNRASDVPLGPNVPMRLAA--SEITLNDGFEVLHDGFLDNLNQGSSVRFVEQDQ	116	
QY	119	THAAIGKNGLEVSVTLPQEWSSLSQSDTEGKNRNVFTGGGGGHPMVTVASDIAEAR	178	
Db	117	KHIVGQNRNVSVTSVLSQEQYARLQSIDPECKDFVFTGGGGAGHAMVTVASDITEAR	176	
QY	179	TRILAKLPDNRHGGKQPKVDVTRSVGVGSASGI-----DDGV--VSEHTSTNNSVRS	230	
Db	177	QRILELLEPKGTG-----ESKGAGESKGVGELRESNGAENTTETQSTSTSSLS	227	
QY	231	DPKFWVSGVCAAGLAGLAATGIAQALATPEDDPTTTPDQAANAASATKQDLTQEA	290	
Db	228	DPKMLAGTVATGVLGLAATGIVQALATPEPDSPTTTPDQAASAATETATRDQLTREA	287	
QY	291	FKNPENQKVNIDANGNAIPSGELXDDIVEQIAQQAQEAQVAVESNAQAQQRYEQ	350	
Db	288	FQNPQKVNIDELGNAIPSGVLKDDVAVIEQAQAQEAQVAVESNAQAQQRYEQ	347	
QY	351	HARQEEQLSSGIGYGLSSALIVAGGIGAGVYTAALHRRNQPAEQTTTTT-----HT	403	
Db	348	QAKRQBELKVSSGAGYGLSGALILGGIGVAVTAALHRRKNQPVQEQTTTTTTTTTSART	407	

404 VVOQGGIPQHKVALMPQERRRFSRRDSQGSVASTHMSDSS-SEVVNPYAEVGGARNS 462
Db VNK PANNTPAQGNVDTPGSDTMESSRRSMASSTSTFTDTSIGVQNPYADV-----K 462
463 LSAHQPEEHYDEVAADPGVSVIQNFSGSPVT-----GRLGTPQGGIQ 507
Db TSLLD-----SQPPTSNSWTSVQNMGNNTDSVYSTIQHPRTDTNGARLLGNPSAGIQ 516
408 STYALDANSGLRLGMGLTSGGETAVSSVNAAPTGPVRFV 549
Db STVARLALSGGLRHDGGLTGGGNSAVNTSNNPPAPGSHRFV 558
RESULT 2
A28182
hemolysin A - Serratia marcescens
C:Species: Serratia marcescens
C:Date: 30-Jun-1989 #sequence_revision 30-Jun-1989 #text_change 08-Oct-1999
C:Accession: A28182
R:Poole, K.; Schiebel, E.; Braun, V.
J. Bacteriol. 170, 3177-3188, 1988
A:Title: Molecular characterization of the hemolysin determinant of Serratia marcescens
A:Reference number: A28182; MUID:88257037
A:Accession: A28182
A:Molecule type: DNA
A:Residues: 1-1608 <POO>
A:Cross-references: GB:M22618; NID:g340726; PIDN:AAA50323.1; PID:g556420
Query Match 5.4%; Score 150.5; DB 2; Length 1608;
Best Local Similarity 20.9%; Pred. No. 0.25;
Matches 122; Conservative 84; Mismatches 236; Indels 141; Gaps 25;
409 33 GTGHLISGTGALGSRSLFSPLRNSMADSVSDSDIPGLTPNPSR-----LAATSETC 84
Db 1008 GSDLTVDAKGGGTQSRSSASQAVTDAAN--GINVNVKDAIYOGTALNGRGKTA 1065
405 85 LIGGPEVLHDKGLDILNTQIGPSAFRVEQA-----DGFH----- 120
Db 1066 VNAGGDRLDAQSDQSESR--SGFNKASAKGGFTADSKNFGAGFGGTHGESSST 1122
406 121 AAIGEKNGLEVSVTLSPQEWSSLSQIDTEGKRNRFVFTGGRGSGHPVTVASDIAEA-RT 179
Db 1123 AQGVNIGS--QGVVELKAGRLTLQGTVDKSGDVSLSAGN-----KVALQAAESTQT 1173
407 180 RLAKLDPNHHGRQPKDVTFRSVGVGSAG--IDGVVYSETHSTTNSVSRSDPKFWVSV 238
Db 1174 RKESLSGNIDLGAGSSDSKEKTGNLSAGGAFDIAKYNESATERQGATIASDGKVTLSA 1233
408 239 ---GATAAGLAGLAATGIAQALALTPEDDPTTTPDQANAA-----ESATKDQLTOEAFK 292
Db 1234 NGKGDALHLQAKVSGGSAAL-----EAKNGGILLESKANEQ----- 1271
409 293 NPEN-----QKVNIDANGNAIP-----SGELXDDIVEIAQAK-----E 327
Db 1272 HKDNWSLGIKAKAGGQTFNKDAGGVDPNTGKDTHTLGLAGLVKVGVEQDKTHTANTGIT 1331
410 328 AGEVARQQAQVESNAQAQRYEDOHARRQBELQLSSGIGYGLSSALIVAGGIGAGVTTALH 387
Db 1332 AGDVTLNSKDFRL-AGARVD-----ADSVQGVKGGDLHVESRDKDVGVKVDVDAGLS 1384
411 388 RRNPQAEOTTTTTTHVVOQQTGGIQQ-----HKVALMPQER-----RRFSRRDSQGS 436
Db 1385 HSNDCSSITSKLSKVGTTPYAGKVKKELEAGVKNVADATTDKYNVARRLPDQDDTTGA 1444
412 437 VASTHMSDSSSEVVNPYAEVGGARNSLSAQPEEHYDEVAADPGVSVIQNFSGSGPVTG 496
Db 1445 VS---FSKAEKGKVTLPATPAG-----EKPGPLWDRGARTVGVGAVKDSITGPAGRG 1493
413 497 RL-----LGTPOGQISTYALLANSGLRLGMGLTSGGETAVSSVN 531
Db 1494 HLKVNADVNNNAVGEQSA---IAGKNGVALQVGGQTQTGTGGE 1533

3
C71513
hypothetical protein CT456 - Chlamydia trachomatis (serotype D, strain UW3/Cx)
C:Species: Chlamydia trachomatis
C:Date: 13-Sep-1998 #sequence_revision 13-Sep-1998 #text_change 08-Oct-1999
C:Accession: C71513
R:Stephens, R.S.; Kalman, S.; Lammel, C.J.; Fan, J.; Marathe, R.; Aravind, L.; Mitche
Science 282, 754-759, 1998
A:Title: Genome sequence of an obligate intracellular pathogen of humans: Chlamydia t
A:Reference number: A71570; MUID:99000809
A:Accession: C71513
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1005 <ARN>
A:Cross-references: GB:AE001319; GB:AE001273; NID:g3328881; PIDN:AAC68056.1; PID:g332
A:Experimental source: serotype D, strain UW-3/Cx
C:Genetics:
A:Gene: CT456
Query Match 5.4%; Score 150; DB 2; Length 1005;
Best Local Similarity 22.8%; Pred. No. 0.14;
Matches 145; Conservative 65; Mismatches 234; Indels 192; Gaps 33;
404 4 GNLGNVNVGNHLLIPAPPPL-PSQFDGAARGGT-----GHLISSTGALGSRSLFSPLR-- 54
Db 431 GNLPTVTIINNKFKTCVAYGPMNSQEAASGYTPSAMRRGRHVDVFGGIFEKANDFNKINWG 490
405 55 -----NSMADSVSD-SDIPGLTPNPSRLAAATSETCILLGFEVLHDKGLDILNTQIGPS 108
Db 491 TQAGPSEDDGIGSFNETPG--AGP---AAPSP-----PSSIPILN----- 528
406 109 AFRVEQADGTHAAIEKNGLEVSVT--LSPQEWSSLSQIDTEGKRNRFVFTGGRGSGHPM 167
Db 529 ---VNVNVGCTNVNIGTGN--VNTTNTTPTQSTDASTDT----- 563
407 168 VTVASDTAEARTRILAKLDPNHHGRQPKDVTFR--SVGVG-----SASGIDDG 214
Db 564 ----SIDDILNT-----NNQDDINTDKDSGAGGVNGDISETESSGDDSG 607
408 215 VVSEHTSTTNSSVSRDPKFWVSV-----GAIAG-----LAG 247
Db 608 SVSSSE-SDKNASVGNDBGPAKMDILSAVRKHLVDVYPGNGSGTEGPLPANQTLGDVID 666
409 248 LAATGIAQALALTEP-----PDDPTTTPDQANAAESAT-----KDOLTOEA----- 290
Db 667 VENKGSADTKLSGNTGAGDDDDPTT---AAVNGAEETILSDTSGIGDDVSDTASSSGD 724
410 291 ---FKNPENOKVNIDANGNAIPSGELXDDIVEIAQAKAEAGEVARQQAQVESNAQAQQR 346
Db 725 ESGGVSSPSSSNKNTAVNGDPSGL---DILAARVRLDKVYPCDNGSGSTEGPLQANQT 781
411 347 YED-----QHARRQBELQLSSGIGYGLSSALIVAGGIGAGVTTALHRRNQPAEQTTTTT 401
Db 782 LGDIVQDMETTGTSTQETVVSF--WKGSTSTESAGGSGVQT---LLFSPPTPTTTTLRT 837
412 402 HT-----VVQOQTGG-IPQHKV--ALMPQERRRFSRRDSQGSVASTHW 442
Db 838 GTGATTTSLMMGGPIKADIITGGGGRIPGGGTLEKLLPRLAHLDISFDAQGDVSTEE 897
413 443 SDSSEVVNPYAEVGGARNSLSAQPEEHYDEVAADPGVSVIQNFSGSGPVTGRLIGTP 502
Db 898 POLGS-IYVNFQETGSGRILA-----FVESAPKPGSA--QVLTGTG-----GDK 940
414 503 GGGIQTSTYALLANSGLRLGMGLTSGGETAVSSVN 538
Db 941 GNLFOAAAATQALGNV-AGKVNLAIOGQKLSSLVN 975
RESULT 4
D85644

hypothetical protein Z1495 [imported] - Escherichia coli (strain O157:H7)
C:Species: Escherichia coli
C>Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 31-Mar-2001
C:Accession: D95644
R:Perina, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
Miller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: D95644
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-2806 <STO>
A:Cross-references: GB:A8005174; NID:g12514354; PIDN:AAG55616.1; GSPDB:GN00145; UWGP:Z14
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: Z1495

Query Match 5.3%; Score 149.5; DB 2; Length 2806;
Best Local Similarity 20.8%; Pred. No. 0.64;
Matches 113; Conservative 88; Mismatches 218; Indels 123; Gaps 27;
QY 22 LPSQTDGAARG--GTGHLISSTGALGSRSLFSPRLNSMADSVDSRDIPGLPNTNPSRLAA 79
DB 68 LPTMGKGFQAQVRGTGEM---ARGLGDAIQSPVKTG-ARILNEFSRMGLP-----GVA 117
QY 80 TSETCLLGGFEVLHDKGPDLDILNTQ-----IGSAFRVEVQADGTHAIGEKNGL 129
DB 118 TVQDIFAG-----SRGADEVITDLPDGNVATDIVKG-----LKATGKAVSDGAKATD 167
QY 130 E-VSVTLSPQEWSSLSQSIDTEKKNRFVFTGGRG---GSGHP-MVT--VASDIAEART- 180
DB 168 EWLTKMSPGAVRALNTPMTEGYNDSNVVAVGNLIGALYPMVAGVARKVGDVTLRK 227
QY 181 -----ILAKLPDNDHGRQPKVDTRSVGVGSAGIDGQVYVSETHFTTNSSVRS 230
DB 228 MLTAGLEKKYIAAGMOPERATALAAEAVDKMPDLFQA-----GLITHSVSAQGS--- 279
QY 231 DKFWVSVCAIAGLAGLAA--TGTAQA-----LALTEPDPTDTPDQAANAESA 281
DB 280 -----AMAAADVLANADYSELAQSPKFOOTFLSDADPOHAQLTD-RQKMDLAKER 329
QY 282 TKDQLTQEAFAKPNOKVNDANGNAIPSGELKDDIVEQIAQOAKE--AGEVARQOAVES 339
DB 330 VADEVRAQLATPELLAVN--AMAAKLGDAQLFNLVTRCTAKTVKSGIVRNATEGAINA 387
QY 340 NQAQOQRYEDQHARRQEEQLQSSGIGY-----GLSSALIVAGGIGA-----GVTTALHRR 389
DB 388 AOGGYSRYQENTALRE-----TAGMGVSPWEGVADATIEGAFAAGMAPPFGAVAGYGR 442
QY 390 NQPAEQTTTTTTHVVQQTGGIPIQ-HKVALMPOER-----RRFSDR-RDSOG 435
DB 443 RQAAEETAMRDATV--QODDAAPQSPESVDPVAAQRESMGNNRQQLQEQYADADMATGE 500
QY 436 SVASTHWSDSSEVNPNPAEYVGGARNSLSAHQPEHIYDEVAADPCYSVIONFSGSGPVT 495
DB 501 DASAHRREAAQSOLL-----NELDEQTKRAVMNELKAPRSELLEYRRLSQKE 550
QY 496 GR 497
DB 551 GR 552

RESULT 5
T34434
hypothetical protein K06A9.1a - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 18-Feb-2000
C:Accession: T34434
R:Geisels, C.; Gattung, S.
submitted to the EMBL Data Library, December 1996
A:Description: The sequence of C. elegans cosmid K06A9.

A:Reference number: Z21525
A:Accession: T34434
A:Status: preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: DNA
A:Residues: 1-2232 <GBI>
A:Cross-references: EMBL:U08046; PIDN:AACT0890.1; GSPDB:GN00028; CESP:K06A9.1a
A:Experimental source: strain Bristol N2; clone K06A9
C:Genetics:
A:Gene: CESP:K06A9.1a
A:Map position: X
A:Introns: 38/1; 75/3; 103/3; 132/2; 158/2; 222/1; 1088/1; 1367/1; 2039/1; 2049/1; 20

Query Match 5.3%; Score 149; DB 2; Length 2232;
Best Local Similarity 19.9%; Pred. No. 0.5;
Matches 112; Conservative 71; Mismatches 238; Indels 142; Gaps 19;
QY 21 PLPSQ--TDGAARGGTGHLISSTGALGSRSLFSPRLNSMADSVDSRDIPG----LPTNPS 74
DB 899 PPSQSTTIGSTGSGTSPGISTT-----SEMTSGQSTQTQPGSTGVTQPS 945
QY 75 RLAAATS--ETCLLGGFEVLHDKGPDLDILNT-----QIGPSAFRVEVQAD 117
DB 946 TVSDTSGSTTVTGSTEGSSSPIPSTQNTNPSTSGSSMSTQTPQSSQSTSPVESFTS 1005
QY 118 GTHAIGEKNGLEVSVTLSPQEWSSLSQSIDTEKKNRFVFTGGRGSGHPMVTVASDIAEA 177
DB 1006 GATSSGSGPTTLTISPSPPSPSTI-----GSSQGSTSPV---STISGQ 1048
QY 178 RFRILAKLPDNDHGRQPKVDTRSVGVGSAGIDGQVYVSETHFTTNSSVRSRDKFWVS 237
DB 1049 STE-----TPGSTGTVTKP-----STVSGSASSGSTATMGSTEASSTSGSSSTSPNPSQS 1099
QY 238 VCAIAGLAGLAAATGIAQALATPEPDDPTTDPQOANA--AESATKDLQLEAFKNPE 295
DB 1100 TSPSTSGATSPGSGGTTLTISPSQSSSTIGSSQGSTSPVYVSTSGDMTSGGTQIPG 1159
QY 296 NOKVNI-----DANGNAIPSGELKDDIVEQIAQOAKE--KEAGEVARQOAVESNAQAQRYE 348
DB 1160 STGTVTQPTSGSGSTSGEITSGQSTQTPRSSISTSPASTSTQQSVSTNSP----- 1213
QY 349 DOHARRQEEQLQSSGIGYGLSALIVAGGIGAGVGTALHRRNOPAQTTTTTTHVVQOQ 408
DB 1214 -----GSTVQTPSTVRGSTSGSTVTTGSTECSST 1243
QY 409 TGGIQHKVALMPOERRRFSDRDSQGSVASTHWSDSSEVNPNPAEYVGGARNSLSAHQP 468
DB 1244 SGS-----SSATLSLSSSPVPSTSQSNP-----STSGSSTPTPNP 1279
QY 469 EHIYDEVAADPCYSVIONFSGGVPVTRGLTGPQGIGQSTYALLANSGLRLGMC---G 525
DB 1280 SOSTSPVYVSTTGTGEMTSHGSTQTPSTIGSTVTP-----STVS-GSNSGSGVTIGSSEA 1333
QY 526 LTRSG--ETAVSSVNAAPTPGPV 546
DB 1334 STSGSFKTSPSISIPVPTSSPI 1356

RESULT 6
T24583
hypothetical protein T06D8.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T24583
R:Palmer, S.
submitted to the EMBL Data Library, April 1995
A:Reference number: Z19909
A:Accession: T24583
A:Status: preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: DNA
A:Residues: 1-1829 <WIL>
A:Cross-references: EMBL:Z49130; PIDN:CAA88964.1; GSPDB:GN00020; CESP:T06D8.1
A:Experimental source: clone T06D8

C:Genetics:
A:Gene: CESP-T06D8.1
A:Map position: 2
A:Introns: 1391/3; 1432/3; 1470/3; 1505/1; 1520/1; 1616/1; 1644/1; 1687/3; 1742/1

Query Match 5.2%; Score 146; DB 2; Length 1829;
Best Local Similarity 21.7%; Pred. No. 0.58;
Matches 118; Conservative 63; Mismatches 256; Indels 108; Gaps 19;

QY 68 GLPTNPSRLAAATSETCLLGFEVLHDKGPLDILNTQIGPSARFEVQADGTHAAIGEN 127
DB 172 GEETTTSAVTEASSEA-----TTTPAGTEASGEET-----TTSATEGS 210
QY 128 GLEVS-----TLSPQWSSLSQIDTE-GKNRFVFTGGRGSGHPMTVASDIAEARTRI 181
DB 211 GEETTVVAVVSSGEEEPASSSTSIPTLSKNDQV-----TEASGEETITAAATEASETTT 266
QY 182 LAKLDPNHHGROPKDYDTRSVGVSAG-----IDGVVSETHSTTNSS 227
DB 267 SAVTEGSGE-----DTTVAVVELSGEQPASSSTSIPTLSKDDQVTEASGEETTTA 318
QY 228 VRSDPKFWVSGVGAAGLAGLAATGIAQALALTPEDDPTTDP-----DOANAESA 281
DB 319 AATEASEBETTTSAVTEG-SGEETTVAVVSSGEEEPASSSTSIPTLSKDDQVTEASGEE 377
QY 282 TKDQLOTFAPKNPENQVN-----IDANG-----NAIPSGELX--DDIVEQ 320
DB 378 TTTAAATEASEBETTTSAVTEGSGEDTTVAVVSSGEEEPASSSTSIPTLSKDDQVTEA 436
QY 321 IAQAKAAGEVARQAVESNAQAQORVEDQ-----HARQEEQLSGGIGYGLSSALIV 374
DB 437 SGEETTTAAATEASEBETTTSAVTEGSGEDTTVAVVSSGEEEPASSSTSIPTLSKDDQV 496
QY 375 AGGICAGVTTALHRRNOPAEQTTTT-----TTHVVOOQTGGIPQHKVALMPQERR 425
DB 497 TEASGEETTTA-AATEASEBETTTSAVTEGSGEETTVAVVSSGEEEPASSSTSIPTLS 554
QY 426 RFSDRDQSGVAST-HWSDSSSEVYNPYAEVGGARNLSAHPHEHYDEVAADPGYSV 484
DB 555 KDDKYTEASGETTTTAAATDASSETTTTSAVTEGSGEETTVAVVSSGEEEPASSSTSI 613
QY 485 IQNFS-----GSGPVNGLIGTPGQIQSTYALLANSGLRLGMGLTSGGETAVSSV 537
DB 614 PRLSKDDQVTEASGEETTTAAATEASEBETTTSAVTEGSGEETTVAVVSSGEEEPASS 673
QY 538 NAAPT 542
DB 674 TS IPT 678

RESULT 7
T13936
collar protein isoform C - fruit fly (Drosophila melanogaster)
C:Species: Drosophila melanogaster
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 17-Nov-2000
C:Accession: T13936
R:McGinnis, N.; Ragnhildstveit, E.; Veraksa, A.; McGinnis, W.
Development 125, 4553-4564, 1998
A:Title: A cap 'n' collar protein isoform contains a selective Hox repressor function.
A:Reference number: 217817; MUID:98453413
A:Status: preliminary
A:Accession: T13936
A:Molecule type: mRNA
A:Residues: 1-1296 <MC>
A:Cross-references: EMBL:AF070064; NID:g3859888; PID:g3859889; PIDN:AAC72898.1
C:Genetics:
A:Gene: cnc
A:Cross-references: FlyBase:FBgn0000338
A:Map position: 3
C:Keywords: leucine zipper

Query Match 5.2%; Score 144.5; DB 2; Length 1296;
Best Local Similarity 19.9%; Pred. No. 0.44;
Matches 149; Conservative 68; Mismatches 213; Indels 319; Gaps 31;

QY 8 NNVNGNHLIPAPPLPSQ-TDGAARGGTGHLISSTGALGSRSLFSLPLNSMADSVDSRDI 66
DB 24 HNTGSSVQTAALQDVQSTSAATGAT--MVVGTG----- 57
QY 67 PGLPNPSRLAAATSETCLLGFEVLHDKGPLDILNTQIGPSARFEVQADGTHAAIGBK 126
DB 58 -CAPTSSGGTSGSA-----LG--EIHIDTASLDPGNANHSPLHPTSELDFTLTHALQDQ 109
QY 127 NGI-----EVSVTLSPQWSSLSQIDTEKN-----RFVFTGGRGSGHP 166
DB 110 RSIWEONLADLYNDLSLOTSPYANLPKDGQPOQPSNSSHLDLSLALLHGFTGGSGAP 169
QY 167 MVTVASDIAEARTILAKLDPNHGRQKQVDVTRSVGVSGASGIDD-----GVV 216
DB 170 LSTAALNDSTPHPRNLGVS--TNNAGR--SDGEEESLYLGRLFGEDEDEYEGELVGVGA 226
QY 217 S--ETHSTTNSSVRS-----DPRFWVSVGAIAAGLAG 247
DB 227 NACEVEGLTTDEFPFGNCFANEVEIGDDEESEAIEVLYKQDYDLGFLDQEI--ING 283
QY 248 LAATGIAQALALTPEDDPT--TTDPD-----QAANAASATKD-----Q 285
DB 284 SYASGNSAATNVKSKPEDETKSSDPSISSESGFKDQDVNAENEAASAASVDDIEKLKALEE 343
QY 286 LTQEAFAKNPENQVINDANGNATP-----SGE-----LXD 315
DB 344 LQDKDKNNENQLEDITNEWNGIPFTIDNETGYIRLPDELNDVILKSEFFLQDDLSN 403
QY 316 DIYEQTAQAQAKEAGEVARQAV-----ESNAQAQORVE----- 348
DB 404 DPVASTSQAAAAFENNOAQRIVSETGEDLLSGEISSKQNRNEAKNDNDPEKADGDSFS 463
QY 349 -----DQHARRQEELQLSS----- 362
DB 464 VSDPEELQNSVGSPFLDLEDKDELQSTVPSYHHPHHPHHPHHPHSHHSHHSMH 523
QY 363 -----GIGYGLSSALIYAGGIGAGVTTALHRRNOPAEQTT 397
DB 524 HAHAAAAAQAQAVQANYGSGVGVG-----GVGVGSGTGSFAQR--QPA----- 573
QY 398 TTTTHTVVQOQTGIPQ-HKVALMPQERRRFSRRDSQGSVASTHWSDSSEVYNPYAEV 456
DB 574 -----AGFHHCHHGRMPRLNRSVSMERLQD---FATYFSPIS--MV 612
QY 457 GGARNLSAHPHEHYDEVAADPGYSVIQNFSGSPVTGRLIGTPGQIQSTYALLANS 516
DB 613 GG-----VSDMSPPYPHY-----PGYSQASPSNGAP-----GTPGQHGQY----- 648
QY 517 GGLRLGMGLTSGGETAVSSVNAAPTGP 545
DB 649 -----GSGANATLQPPPPPP 664

RESULT 8
S04722
puff 74E protein - fruit fly (Drosophila melanogaster)
N:Alternate names: ecdysone-induced protein E74B; ets-related protein E74B
C:Species: Drosophila melanogaster
C:Date: 28-Feb-1990 #sequence_revision 28-Feb-1990 #text_change 21-Jul-2000
C:Accession: S04722; B34692
R:Janknecht, R.; Taube, W.; Luedicke, H.J.; Pongs, O.
Nucleic Acids Res. 17, 4455-4464, 1989
A:Title: Characterization of a putative transcription factor gene expressed in the 20
A:Reference number: S04722; MUID:89315191
A:Accession: S04722
A:Molecule type: DNA
A:Residues: 1-883 <JAN>
A:Cross-references: EMBL:X15087; NID:g7513; PIDN:CAA33195.1; PID:g7514
A:Note: Gln-867 was also found

C:Genetics:
A:Gene: SPDB:SPAC89.04
A:Map position: 1

Query Match 5.1%; Score 141.5; DB 2; Length 647;
Best Local Similarity 19.0%; Pred. No. 0.25;
Matches 103; Conservative 64; Mismatches 207; Indels 169; Gaps 20;

QY 114 VQADGTHAAIGKNGLEVSTLSPOEWSLQSIDTEKKNREFTVGTGRGGSGHPMTVVASD 173
DB 94 LQEPGGTQIVKDAQVDEPLEPIASSALGTVEPTDNK-----PSASTSTA 141
QY 174 IABARTILAKLPDNGHGQPKVDVTRSVGVSASGIDGVVSEHTSTTN----- 225
DB 142 VPTTEARTSITEFANSPSSSSSSASTKSTATTQSA---DYVVAEHFAPORNDQELGNSP 198
QY 226 SSVSRDPKFWVSV---CAIAAGLAGLAATGIAQALALTPP---PDDP-----TTTDPD 272
DB 199 ASITSKPAATSAQPPSSKVENMAKATSQPIITAEKEIPELKPPEPEAIMSKEINTHD 258
QY 273 QAA-----NAAESATKQLTQAEAKN-----P 294
DB 259 QAAATTTAAVASASTTATAESHAVADGIMDNVLESIGENVQOETVFEFEDASDIPHADVIP 318
QY 295 ENQKVNID-----ANGNAIPSG---ELXDDIVEQIAQAQAKEAGEV 331
DB 319 HTTTVTVEESPIALGOGVYTHEATTSARASASGIPGAFEEVQOTVOEDLPHPTAEIVEI 378
QY 332 ARQAVESNAQAQORYEDOHARRQEELQSGSIGYGLSSAL-----IV 374
DB 379 ARFAEQPVRAQOPEYESSVV--QEAETVTDVKGVSSTVKNEVNPSTIPTSESNPVA 436
QY 375 AGGIGAGVTTALHRRNQ-----PAE-----OTTTHTVVVOQTGGIPQHKVAMQER 424
DB 437 VGG-----TTAEHPVQEAHTAPTETAHDFSKETTTASKRVSKHDKASAEKHVKARKP--- 488
QY 425 RFRDRRDSQSVASTHWSDSSESVNVNPAEFGGARNLSAQHPPEEHIYDEVAADPGYSV 484
DB 489 -----SSGQEPPTSTPAKSNOSKHHARRSKQ-----ASAP----- 521
QY 485 IONFSGSGPVTRGLIGTPQGIQSTYALLANSGLGLRMGGLTSGGETAVASVNA-----P 541
DB 522 ----SSPGTTSAAVPGKKSAIEAA-ADPTSDATVSKHAAGSGSATTPSPGSATTKP 576
QY 542 TPG 544
DB 577 TPG 579

RESULT 11
A55182
aggrecan precursor - mouse
C:Species: Mus musculus (house mouse)
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 21-Jan-2000
C:Accession: A55182; S55329; S50207; S51355; I78532; I58123
R:Walcz, E.; Deak, F.; Erhardt, P.; Coulter, S.N.; Fueleop, C.; Horvath, P.; Doege, K.J.
Genomics 22, 364-371, 1994
A:Title: Complete coding sequence, deduced primary structure, chromosomal localization,
A:Reference number: A55182; MUID:95104847
A:Accession: A55182
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: mRNA
A:Residues: 1-2132 <WAL>
R:Watanabe, H.; Gao, L.; Sugiyama, S.; Doege, K.; Kimata, K.; Yamada, Y.
Biochem. J. 308, 433-440, 1995
A:Title: Mouse aggrecan, a large cartilage proteoglycan: protein sequence, gene structure
A:Reference number: S55329; MUID:95289972
A:Accession: S55329
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-567, 'G', 569-1728, 'I', 1730-2132 <WAT1>

A:Cross-references: GB:U22901; NID:g886014
R:Glumoff, V.; Savontaus, M.; Vehanen, J.; Vuorio, E.
Biochim. Biophys. Acta 1219, 613-622, 1994
A:Title: Analysis of aggrecan and tenascin gene expression in mouse skeletal tissues
A:Reference number: S50206; MUID:95035091
A:Accession: S50207
A:Molecule type: mRNA
A:Residues: 350-481, 'R', 483-506 <GLU1>
A:Cross-references: EMBL:X80279; NID:g673432
R:Glumoff, V.
submitted to the EMBL Data Library, July 1994
A:Reference number: S51355
A:Accession: S51355
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 350-383, 'CPVMSQSRPMAA' <GLU2>
A:Cross-references: EMBL:X80279
R:Watanabe, H.; Kinata, K.; Line, S.; Strong, D.; Gao, L.Y.; Kozak, C.A.; Yamada, Y.
Nature Genet. 7, 154-157, 1994
A:Title: Mouse cartilage matrix deficiency (cmd) caused by a 7 bp deletion in the agg
A:Reference number: I58123; MUID:95004579
A:Accession: I78532
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 211-326 <WAT2>
A:Cross-references: GB:S73722; NID:g765215; PIDN:AAB32160.1; PID:g765216
A:Accession: I58123
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 211-240, 'MCTASLRWRVRSFMRHPQNRSPSRQPT', 'AGGWHANPQASSTWPGRAVWTCALAGW'
A:Cross-references: GB:S73720; NID:g765211; PIDN:AAB32159.1; PID:g765212
C:Genetics:
A:Map position: 7
A:Introns: 253/1
C:Superfamily: aggrecan; C-type lectin homology; complement factor H repeat homology;
C:Keywords: cartilage; extracellular matrix
F:1-19/Domain: signal sequence #status predicted <SIG>
F:44-135/Domain: immunoglobulin homology <IMM>
F:170-247/Domain: link protein repeat homology <LNK1>
F:268-349/Domain: link protein repeat homology <LNK2>
F:504-581/Domain: link protein repeat homology <LNK3>
F:602-683/Domain: link protein repeat homology <LNK4>
F:1922-2043/Domain: C-type lectin homology <LCH>
F:2049-2105/Domain: complement factor H repeat homology <FHD>

Query Match 5.0%; Score 140; DB 1; Length 2132;
Best Local Similarity 20.5%; Pred. No. 1.7;
Matches 129; Conservative 81; Mismatches 241; Indels 178; Gaps 29;

QY 4 GNIGNNVNHLIPAPPPLPSQTDGAARGGTGHLITSTGTGALGSRSLFSPLRNSMADSVDS 63
DB 906 GDLSGLPSGGGTEITSTSGAETSGLPSGGD--LETSTSGVDDVSGIPTGRELETSSAG 964
QY 64 -RDIPGLTNP-SRLAAATSETCLGGFVLHDKGLDILNTQIGPSAFRVEVQADGTHAA 122
DB 965 VEDLSGLPSGGESETSTSG-----GIE-----DISLPTG-GES-----LETSSAG----- 1004
QY 123 ICEKNGLEVSTLSPOEWSLQSIDTEKKNREFTVGTGRGGSGHPMTVVAS---DIAEART 179
DB 1005 VGDLSGL-----PSGESLETSSAGAEVDVQLPTERGG---LETSSAGVEDITVLP 1053
QY 180 RILAKLDPNHHGGRPKVDVTRSVGVSASGIDGVDVSETHSTSTSSVRSRDPKFWVSVG 239
DB 1054 -----GRE--SLETSSAGVEDVSGLPSPG---REGLETSSAGIEDISVPFTEAE 1096
QY 240 AIAAGLAGLAATGIAQALALTPEDDPTTDP-----DOAANAES--- 280
DB 1097 GLDTSASGGYVSGI-----PSGGDGTSTASGVEDVSGLPSSGEGLETSSAGVEDLGP 1149
QY 281 ATKDQLTQAEAFKNPENQKYNIDANGNAIPSG-----LXDDIVEQIAQAQAKEAGEV 331
DB 1150 STRDSL-----EYSASGVDDVTG---FSGRCDPSTSVSGVGDGF-----SGLP 1189


```
QY 332 ARQAVESNAQAORYEDQHARROEELQLSSG----- 363
      : : : : : : : : : : : : : : : : : :
Db 1190 SKEGLETASGAGNEDLSGLPSGKEDLVGSAGALDFCKLPPTGLSGQTPVNGFFSPGS 1249
      : : : : : : : : : : : : : : : : : :
QY 364 -----IGYGLSSALIVAGGIGAGYTTALHRRNPAAEQTTTTTHVVVQQQTGGIP-QH 415
      : : : : : : : : : : : : : : : : : :
Db 1250 GEYSGADIGSGPSGLPDFSGLPSGFTVSLVDSTLVEVITATTSLEBGRGTIGISGSG 1309
      : : : : : : : : : : : : : : : : : :
QY 416 KVALMPQERRRFSDRDSOGSVASTHWSDSSEVVNPYAEVGGARNSLSAHQPEEHIYDE 475
      : : : : : : : : : : : : : : : : : :
Db 1310 EVSGLP--LGELDSSADISGLPSGTELSGASGSPDSSGETSG-----FFD 1353
      : : : : : : : : : : : : : : : : : :
QY 476 VAADP--GYSVIONFSGSGPVTGRLTGTCQGIQSTYALL-----ANSGLRLG 522
      : : : : : : : : : : : : : : : : : :
Db 1354 VSQPTFGSSCVSEETSGEISQPSGTPDTTATSGVTNELNLSGQPDVSGDGSGLILG 1413
      : : : : : : : : : : : : : : : : : :
QY 523 MG---GLTS--GGETA--VSSVNAAPTGPV 546
      : : : : : : : : : : : : : : : : : :
Db 1414 SGSSGITSVSGTSGTSDLSGQPSGFPV 1442
      : : : : : : : : : : : : : : : : : :

RESULT 12
A26039
Iga-specific metalloendopeptidase (EC 3.4.24.13) precursor - Neisseria gonorrhoeae (stra
N:Alternate names: Iga protease; immunoglobulin A1 proteinase
C:Species: Neisseria gonorrhoeae
A:Variety: strain MS11
C:Date: 05-Oct-1988 #sequence_revision 05-Oct-1988 #text_change 08-Dec-2000
C:Accession: A26039; S09386
R:Pohlner, J.; Halter, R.; Beyreuther, K.; Meyer, T.F.
Nature 325, 458-462, 1987
A:Title: Gene structure and extracellular secretion of Neisseria gonorrhoeae Iga proteas
A:Reference number: A26039; MUID:87115823
A:Accession: A26039
A:Molecule type: DNA
A:Residues: 1-1532 <POH>
A:Cross-references: GB:X04835; NID:g44868; PIDN:CAA28538.1; PID:g44869
A:Note: the authors translated the codon AAG for residue 668 as Asn
R:Halter, R.; Pohlner, J.; Meyer, T.F.
EMBO J. 8, 2737-2744, 1989
A:Title: Mosaic-like organization of Iga protease genes in Neisseria gonorrhoeae generat
A:Reference number: S09386; MUID:90060036
A:Accession: S09386
A:Status: not compared with conceptual translation
A:Molecule type: DNA
A:Residues: 281-325, 'N', 327-337, 'N', 339-427, 'M', 429-531, 'N', 533-615, 'V', 617-631, 'N', 633-
A:Experimental source: strain MS11
C:Genetics:
A:Gene: Iga
C:Superfamily: Iga-specific metalloendopeptidase
C:Keywords: hydrolase; metalloproteinase; transmembrane protein
F:1-27/Domain: signal sequence #status predicted <SIG>
F:28-1532/Product: immunoglobulin A1 proteinase #status predicted <MAT>
F:986-987/Cleavage site: Pro-Ser (autolytic) #status predicted
F:1018-1019/Cleavage site: Pro-Ser (autolytic) #status predicted
F:1121-1122/Cleavage site: Pro-Ala (autolytic) #status predicted

Query Match 4.9%; Score 137; DB 2; Length 1532;
Best Local Similarity 21.0%; Pred. No. 1.6;
Matches 102; Conservative 60; Mismatches 183; Indels 140; Gaps 20;

QY 34 TGHLSSTGALGSRSLFSPLRNSMADSVDRDIPGLTPNPSRLAAATSETCILLGGFEVLH 93
      : : : : : : : : : : : : : : : : : :
Db 782 TGYVTCNTGNLSKAL-----NSFDRTRNGNVNLNQNAALVLGKAAALWGKIQ--- 829
      : : : : : : : : : : : : : : : : : :
QY 94 DKGPLDLNTQIGPSAFRVEQADGTHAAIGKNLEVSVTLSQPQWSSLSQSDTTEKNR 153
      : : : : : : : : : : : : : : : : : :
Db 830 -----CQGNRSVSLNQHSHWLTGDSQVHNLSDSHIHLNNASDAQSANKYH 877
      : : : : : : : : : : : : : : : : : :
QY 154 FVFTGGRGGSG--HPMVTVASDIAEARTRILAKLDPDNHGGROPKPDYDTRSVGVGSASGI 211
      : : : : : : : : : : : : : : : : : :
```

```
Db 878 TIKINHLSGNGHPHYTLDLAKNLGD---KVLVK---ESASGHYQLHVQNKT-GEPNQEG 930
QY 212 D--DGVVSETHTTSSSVSRDPKFWVSYCAIAAGLAGLAAT-----GIAQALALPEPD 264
      : : : : : : : : : : : : : : : : : :
Db 931 DLFDA-----SSVQDRSLFVSLANHYVDLGAALRYTIKTENGTRILNPNYAGNG 979
      : : : : : : : : : : : : : : : : : :
QY 265 DPTTTDDQAAANAESATKDQLTQEA-FKNPENOKYNIDANGNAIPSGELXLDIVQIAQ 323
      : : : : : : : : : : : : : : : : : :
Db 980 RPVKPAPSPAANTASQAOKATQTDGAIAKPNIV-----APPSQANQAEALRQ 1031
      : : : : : : : : : : : : : : : : : :
QY 324 QAKEAGEVARQQAQVESAQAQRYED-----OHARRQBELQLSSGIGYGLSALIVAGG 377
      : : : : : : : : : : : : : : : : : :
Db 1032 QAK-ABQVRKQAQAEAKVARQKDEEAKKAAEIAEQEEAR-----KAEELAAKQ 1081
      : : : : : : : : : : : : : : : : : :
QY 378 IGAGVTTALHRRNQPAEQTTTTTHVVQQQTGGIPQHKVALMPQERRR----- 426
      : : : : : : : : : : : : : : : : : :
Db 1082 KAAEKAKARELAKQAEAS-----HQNAPKPKRRRRRAILPRPAPV 1124
      : : : : : : : : : : : : : : : : : :
QY 427 FS---DRRDSQGSVASTHWSDS-----SEVNPYAEVGGARNSLSAHQPEE 470
      : : : : : : : : : : : : : : : : : :
Db 1125 FSLDDYDAKDN-----SESSIGNLARVIPRMGRELINDYEEI-----PLE 1164
      : : : : : : : : : : : : : : : : : :

QY 471 HIYDE 475
      : :
Db 1165 ELEDE 1169
      : :

RESULT 13
A83412
hypothetical protein PA1874 [imported] - Pseudomonas aeruginosa (strain PA01)
C:Species: Pseudomonas aeruginosa
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: A83412
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrener, P.; Hickey, M.J.; L
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; L
.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pa
A:Reference number: A82950; MUID:20437337
A:Accession: A83412
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-2468 <STO>
A:Cross-references: GB:AE004613; GB:AE004091; NID:g9947856; PIDN:AAG05263.1; GSPDB:GN
A:Experimental source: strain PA01
C:Genetics:
A:Gene: PA1874

Query Match 4.9%; Score 137; DB 2; Length 2468;
Best Local Similarity 21.5%; Pred. No. 3.2;
Matches 135; Conservative 57; Mismatches 275; Indels 160; Gaps 28;

QY 2 PIGNLGNVNGNHLIPPAPLPQSQT-----DGAARGGTGHLISSTGALGSRSLFPLR 54
      : : : : : : : : : : : : : : : : : :
Db 455 PIGQTSADANGNNSFTPGSQLPDCTVVNVYARDAAGNSSPATSTITVDGVAPNAPVVEPSN 514
      : : : : : : : : : : : : : : : : : :
QY 55 NS-----MADSVDSRDIPGLTPNPSRLAA-----ATSETCLGGFEVLHDKGPLDILN 102
      : : : : : : : : : : : : : : : : : :
Db 515 GSELGSTAEPGSSVTLTDGNGNPIGOTTADANGNWSFTPTPLPDGTVV--NVVARDAAG 572
      : : : : : : : : : : : : : : : : : :
QY 103 TQIGPSAFRVEQADGTHAAIGKNLEVSVTLSQPQWSSLSQSDTTEKNRF 154
      : : : : : : : : : : : : : : : : : :
Db 573 NSFPASVTVDAVAPAT-PTVDPNSNGTTLTSGTAEPGSSVTLTDGNGNPIQGVTDAGSGNW 631
      : : : : : : : : : : : : : : : : : :
QY 155 VFTGGR-----GGSGHPMTVVASDIAEARTRILAKLDPDN-----HGRQRP 195
      : : : : : : : : : : : : : : : : : :
Db 632 TETPSTPLPNGTVVNVATATDPGNSASSPASVTVDAVAPATPV---VNPSTGTTLSGTAE 688
      : : : : : : : : : : : : : : : : : :
QY 196 KDYDTRSVGVGSAG--IDGVVSEYHTTIT---NSSVSRSDPKFWVSYCAIAAGLAGLAA 250
      : : : : : : : : : : : : : : : : : :
Db 689 GATVTLTDGNGNPIQGVTDAGSGNWSFTPTPLPNGTV-----VNATADTASGNTS 739
      : : : : : : : : : : : : : : : : : :
```


QY 483 SVIQFSGGPVTGR----LIGTPGGIQSTYALLANSGLRLGMG 524
|: || | : | : || | : : | || | : : |
Db 1189 SM-----GSMPTSFPRGELVGTAAQEKRTILVENAPSGYLKISSG 1229

Search completed: September 27, 2001, 14:21:42
Job time: 78 sec

THIS PAGE BLANK (USPTO)

Db 292 DVNEFDQYLPNGHPGVPATHQGVTTYSGTGYGISSASSPAGAGHAMMAKQPPQPPA 351
QY 269 TDPOAANAASATKDQLTQEAFAKNPENQKVNIDANGNAIPSGELXDDIIVEQIAQAQKEA 328
Db 352 QPPAQ-----HTLPS-----TEREQQPQQR 372
QY 329 GEVARQAVESNAQAQORYEDQHARROEELQLSSGIG-----YGLSSALIVAGGIGAV 382
Db 373 PHIKTEQLSPSHNSEOOHPQOOQQOQL-----GYGSFNLQHYGFYPPPI-----419
QY 383 TTALHRRNQPAEQTTTTTHTVVQQQGGIPQHKVQALMPQERRRFSRRSQGSVAS--T 440
Db 420 -----TRSEYDTEHQNSG-----SYISHAAGOSGSLYSTFT 451
QY 441 HWSDSSEVVNPNYAEGVGARNLSAHOPEEH 471
Db 452 YMNPTQRPWYTPIDATSGVPTIPQTHSPQQH 482

RESULT 2
HLIA_SERMA
ID HLIA_SERMA STANDARD; PRT; 1608 AA.
AC P15320;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 01-NOV-1990 (Rel. 16, Last annotation update)
DE HEMOLYSIN PRECURSOR.
CN SHLA

OS Serratia marcescens.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Serratia.
OX NCBI_TaxID=615;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 31-40.
RC STRAIN-SN8;
RX MEDLINE=86257037; PubMed=3290200;

RA Poole K., Schiebel E., Braun V.;
RT "Molecular characterization of the hemolysin determinant of Serratia marcescens";
RL J. Bacteriol. 170:3177-3188(1988).

CC -!- FUNCTION: BACTERIAL HEMOLYSINS ARE EXOTOXINS THAT ATTACK BLOOD CELL MEMBRANES AND CAUSE CELL RUPTURE BY MECHANISMS NOT CLEARLY DEFINED.
CC -!- FUNCTION: CELL-BOUND HEMOLYSIN, WHICH RELEASES HEME-IRON FROM ERYTHROCYTES BY INTERACTION WITH THE ERYTHROCYTE MEMBRANE. SHLA REQUIRES SHLB FUNCTION.
CC -!- SUBCELLULAR LOCATION: OUTER MEMBRANE.
CC -!- SIMILARITY: TO P.MIRABILIS HEMOLYSIN (HPMA).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@sib-sib.ch).

CC -----
DR EMBL; M22618; AAA50323.1; .
DR PIR; A28182;
KW Hemolysis; Toxin; Outer membrane; Signal.
FT SIGNAL 1 30
FT CHAIN 31 1608 HEMOLYSIN.
SQ SEQUENCE 1608 AA; 165078 MW; D669B476FE7DAD51 CRC64;

Query Match 5.4%; Score 150.5; DB 1; Length 1608;
Best Local Similarity 20.9%; Pred. No. 0.42; 236; Indels 141; Gaps 25;
Matches 122; Conservative 84; Mismatches

QY 33 GPGHLISSTGALGSRLSPRLNSMADSVDSRDIPGLPTNPSPR-----LAAATSETC 84
Db 1008 GSDLIVDAKGEGETORSNSASQAVTGSIDAN--GINVNVKKDAIYQGTALNGRGRKTA 1065

QY 85 LLGGEVFLHDKGLDILNTQIGPSAFREVOA-----DQTH-----120
Db 1066 VNAGGDRLRLDQADKQSESR---SGFNKASAKGGFTADSKNFCAGFGGTHGESSST 1122
QY 121 AATGKNGLEVSYTLSPQWSSLSQSDTEGKNRFVFTGGRGSGHPMVTVASDAEA-RT 179
Db 1123 AQVNTSG--QOGVELKAGRLDTLQGTDKSQGDVSLAGN-----KVALQAESTQT 1173
QY 180 RILAKLDPDNHHGRQPKDVTDRSVGVSASG--IDDGVVSETHSTSTNNSVSRDPKFWVS 238
Db 1174 RKESKLSGNDLIDAGSSDSKEKTGGLNLSAGAFDIAKNVESATEROGATIASDGKVTLSA 1233
QY 239 ---GATAAGLAGLAATGIAQALALTPEPDDPTTTPDQANAA---ESATKDQLTQEAfk 292
Db 1234 NGKGDALHLQGAQKVGSGSAAL-----EAKNGGILLISAKNEQ-----1271
QY 293 NPEN-----OKVNIDANGNAIP-----SGELXDDIIVEQIAQAQAK-----E 327
Db 1272 HKDNWSLGIKANKAGGQTFNKGKVDPTNGRDTHTLGLAGLVGVEQDKYTHANTGIT 1331
QY 328 AGEVARQAVESNAQAQORYEDQHARROEELQLSSGIGYGLSSALIVAGGIGACTTALH 387
Db 1332 AGDVTLSNGKDTL-AGARVD-----ADSVQGVKVGDLHVESRKDVENGKVDVDAGLS 1384
QY 388 RNQPAEQTTTTTHTVVQQQGGIPQ-----HKVALMPQER-----RRESDRDSQGS 436
Db 1385 HSNPDGSSITSKLSKVGTPRYAKVKEKLEAGVNVKADATTDKYNVNSVARRLDPODDTTGA 1444
QY 437 VASTHWSDSSEVVNPNYAEGVGARNLSAHOPEEHIDEVAADPGYSVIONFSGSGPVTG 496
Db 1445 VS---FSAEGKVTLPATPAG-----EKPGPLMDRGARTVGGAVKDSITGPAGROG 1493
QY 497 RL-----IGTPGOGIOSTYALLANSGLRLGMLGLT--SGGE 531
Db 1494 HLKVNADVWNNNAVGEQSA---IAGKVALQVGGQTQTGTGE 1533

RESULT 3

E74B_DROME
ID E74B_DROME STANDARD; PRT; 883 AA.
AC P11536; Q9VVI7;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE ECDYSONE-INDUCED PROTEIN 74EFB (ETS-RELATED PROTEIN E74B).
GN EIP74EF OR E74B OR CG6285.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89315191; PubMed=2501755;
RA Janknecht R., Taube W., Lueddecke H.-J., Pongs O.;
RT "Characterization of a putative transcription factor gene expressed in the 20-OH-ecdysone inducible puff 74EF in Drosophila melanogaster";
RL Nucleic Acids Res. 17:4455-4464(1989).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=90199900; PubMed=2107982;
RA Burtis K.C., Thummel C.S., Jones C.W., Karim F.D., Hogness D.S.;
RT "The Drosophila 74EF early puff contains E74, a complex ecdysone-inducible gene that encodes two ets-related proteins.";
RL Cell 61:85-99(1990).
RN [3]
RP SEQUENCE FROM N.A.
RX STRAIN=BERKELEY;
RC MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D., Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F., George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,

RA Sutton G.G., Wortman J.R., Vandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champagne M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA April J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,
RA Bailew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Bottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreria S., Fleischmann W.,
RA Fossier C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hustin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lako P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.D.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskis R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RA "The genome sequence of *Drosophila melanogaster*.";
RL Science 287:2185-2195(2000).
CC -!- SUBCELLULAR LOCATION: NUCLEAR.
CC -!- ALTERNATIVE PRODUCTS: THE E1P74EF ENCODES FOR TWO PROTEINS THAT
CC DIFFER IN THEIR N-TERMINAL SECTION.
CC -!- INDUCTION: THE EXPRESSION OF THIS PROTEIN IS DEVELOPMENTALLY
CC REGULATED AND IS CORRELATED WITH THE 20-OH-ECDYSONE INDUCED
CC ACTIVITY OF PUFF 74EF.
CC -!- SIMILARITY: BELONGS TO THE ETS FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X15087; CA33195.1; -;
DR EMBL; M37083; AAA28494.1; -;
DR EMBL; AE003523; AAF49324.1; -;
DR PIR; S04722; S04722.
DR PIR; B34692; B34692.
DR HSP; P04002; 1ATF.
DR TRANSFAC; T00210; -;
DR FlyBase; FBgn0000567; E1p74EF.
DR InterPro; IPR000418; -;
DR Pfam; PF00178; Ets; 1.
DR PRINTS; PR00454; ETSDOMAIN
DR PROSITE; PS00345; ETS_DOMAIN_1; 1.
DR PROSITE; PS00346; ETS_DOMAIN_2; 1.
DR PROSITE; PS00061; ETS_DOMAIN_3; 1.
KW Nuclear protein; Transcription regulation; DNA-binding; Polymorphism.
KW Alternative splicing; Developmental protein; (ACIDIC).
FT DOMAIN 78 280 ASP/GLU-RICH (ACIDIC).
FT DOMAIN 281 759 ALA/GLN/SER-RICH.
FT DOMAIN 760 883 ARG/LYS-RICH (BASIC).
FT DNA_BIND 787 869 ETS-DOMAIN.
FT DOMAIN 27 51 POLY-SER.
FT DOMAIN 72 79 POLY-ALA.

FT	DOMAIN	289	299	POLY-ALA.
FT	DOMAIN	372	378	POLY-GLN.
FT	DOMAIN	406	415	POLY-GLN.
FT	DOMAIN	434	437	POLY-SER.
FT	DOMAIN	486	489	POLY-ALA.
FT	DOMAIN	510	524	POLY-GLN.
FT	DOMAIN	614	630	POLY-GLN.
FT	DOMAIN	639	642	POLY-GLN.
FT	DOMAIN	651	661	POLY-ALA.
FT	DOMAIN	695	703	POLY-SER.
FT	DOMAIN	707	711	POLY-ALA.
FT	DOMAIN	712	733	POLY-ALA.
FT	VARIANT	867	867	H -> Q
SQ	SEQUENCE	883 AA;	94820 MW;	148D5031A18D1409 CRC64;

Query Match	5.1%	Score 144;	DB 1;	Length 883;
Best Local Similarity	20.8%	Pred. No. 0.46;		
Matches 124;	Conservative 68;	Mismatches 197;	Indels 208;	Gaps 26;
QY	39	SSGTALGSRSLFSLRNMSADSVDSRDPLGLTNPFSRLAAATSETCLLGGFEVLHD----	94	
Db	38	SSLSLSLSSSSSSSLSSATPTPVAS---PVTPTSPPPAAAAPAEASPPAGAELOEDGQQA	94	
QY	95	---KGPL---DILANTQIGPSAFRVEVQADG---THAIGEKNGLEVSVTLSPOEWSSL	143	
Db	95	KTQEDPTMKDQMLEKTROEVKDPNVNEPFGAIVDTESVMAKOSPSPVASTKVP-----	151	
QY	144	QSIDTEGNRFVFTGGGSGHPMV-----TVASDIAEARTILAKLDPNHGGROP	195	
Db	152	EFISNKS-----PPVQDEESESVASDCREFKVLNHLRQQQHH--HSP	194	
QY	196	KVDV--TRSVGVSASGIDD--GVVSETHSTNTSSVSDPKFVSVGTAAGLAGLAATG	252	
Db	195	SSPDKTRST-----LDDVSKILMERKQOLQRSSV-----	223	
QY	253	IAQALALTP-----PDDPTTTDPQAAANAESAATKDOLTOFAFKNPENOKVNIDA	303	
Db	224	IVAAPTLQOQHQPMDSIEDIEDVDADADVEDADELELQYQYQYQYQYQYQYQYQYQYQY	283	
QY	304	NG-----NAIP-----SGELXDDIV--EQIAQQAQKEAGEVARQQAVE----	338	
Db	284	GCATSAASAAAAAASVRRRTYSGTESDSDSAQCERARMRLKPERKAERSAAYKKSML	343	
QY	339	-----SNAQAQRYEDQHARRQBELQSSGIGYGLSSALLVAGGIGAGVTT	384	
Db	344	KRYTEIPTVKQSTSPAPQOQLQOQHLLQOQOQOQPHNGSTF-----AGATA	390	
QY	385	ALHRRNQPAEOTTTTTHVVOQGTGIPQHKVALMPQERRRFRSDRDSQGSVASTHWS	444	
Db	391	LLHUK---TEQNTLITPLQOQOQ-----QQOQOQ-----	416	
QY	445	SSSEVVNPYAEVGGARNSLSAHPBEHIYDEVAADPGYVQNFSGSGPV-----	494	
Db	417	-----LHGAAGNGSSNGNNAHQOQOPL-----AIPQRPDLLHLLSGGAIHNP	466	
QY	495	---TGLRIGTPG--OGIOSTYALLANSQG-----LRLQMGGLTSGGETAVSVNAA	540	
Db	467	ATTGSGFPSPADSGVSDVDS--SSSGQGPACDELKARLGMPPATSAASAAAAA	521	

RESULT	4
PGCA_MOUSE	STANDARD;
ID	PGCA_MOUSE
AC	Q61282; Q64021;
DT	01-NOV-1997 (Rel. 35, Created)
DT	01-NOV-1997 (Rel. 35, Last sequence update)
DT	01-NOV-1997 (Rel. 35, Last annotation update)
DE	AGGRECAN CORE PROTEIN PRECURSOR (CARTILAGE-SPECIFIC PROTEOGLYCAN CORE
DE	PROTEIN) (CSCP).
GN	AGC1 OR AGC.
OS	Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BALB/C; TISSUE=Cartilage;
 RX MEDLINE=95104847; PubMed=7806222;
 RA Walcz E., Deak F., Erhardt P., Coulter S.N., Fueleop C., Horvath P.,
 RA Dege K.J., Glant T.T.;
 RT "Complete coding sequence, deduced primary structure, chromosomal
 RT localization, and structural analysis of murine aggrecan.";
 RL Genomics 22:364-371(1994).
 RN [2]
 RP SEQUENCE OF 211-326 FROM N.A.
 RC STRAIN=129/SV;
 RX MEDLINE=95004579; PubMed=7920633;
 RA Watanabe H., Kimata K., Line S., Strong D., Gao L.-Y., Kozak C.A.,
 RA Yamada Y.;
 RT "Mouse cartilage matrix deficiency (cmd) caused by a 7 bp deletion in
 RT the aggrecan gene.";
 RL Nat. Genet. 7:154-157(1994).
 CC -1- FUNCTION: THIS PROTEOGLYCAN IS A MAJOR COMPONENT OF EXTRACELLULAR
 CC MATRIX OF CARTILAGINOUS TISSUES. A MAJOR FUNCTION OF THIS PROTEIN
 CC IS TO RESIST COMPRESSION IN CARTILAGE. IT BINDS AVIDLY TO
 CC HYALURONIC ACID VIA AN AMINO-TERMINAL GLOBULAR REGION. MAY PLAY A
 CC REGULATORY ROLE IN THE MATRIX ASSEMBLY OF THE CARTILAGE.
 CC -1- SUBCELLULAR LOCATION: SECRETED; EXTRACELLULAR MATRIX (BY
 CC SIMILARITY).
 CC -1- DOMAIN: TWO GLOBULAR DOMAINS, G1 AND G2, COMPRISE THE AMINO
 CC TERMINUS OF THE PROTEOGLYCAN, WHILE ANOTHER GLOBULAR REGION, G3,
 CC MAKES UP THE COOH TERMINUS. G1 CONTAINS LINK DOMAINS AND THUS
 CC CONSISTS OF THREE DISULFIDE-BONDED LOOP STRUCTURES DESIGNATED AS
 CC THE A, B, B' MOTIFS. G2 IS SIMILAR TO G1. THE KERATAN SULFATE (KS)
 CC AND THE CHONDROITIN SULFATE (CS) ATTACHMENT DOMAINS LIE BETWEEN G2
 CC AND G3.
 CC -1- PTM: CONTAINS MOSTLY CHONDROITIN SULFATE, BUT ALSO KERATAN SULFATE
 CC CHAINS, N-LINKED AND O-LINKED OLIGOSACCHARIDES.
 CC -1- DISEASE: DEFECTS IN AGC1 ARE THE CAUSE OF CARTILAGE MATRIX
 CC DEFICIENCY (CMD), AN AUTOSOMAL RECESSIVE SYNDROME CHARACTERIZED BY
 CC CLEFT PALATE, SHORT LIMBS, TAIL AND SNOUT. MUTATION IN STRAIN CMD
 CC CAUSES ABSENCE OF AGGREGAN BY TRUNCATION OF THE PROTEIN (MUTATION
 CC IN THE G1 DOMAIN).
 CC -1- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE V-TYPE DOMAIN.
 CC -1- SIMILARITY: CONTAINS 4 LINK DOMAINS.
 CC -1- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
 CC -1- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.
 CC -1- SIMILARITY: CONTAINS 1 SUSHI (SCR) REPEAT.
 CC -1- SIMILARITY: BELONGS TO THE AGGREGAN/VERSICAN PROTEOGLYCAN FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; L07049; AAC37670.1; -
 CC EMBL; S73722; AAB32160.1; -
 CC EMBL; S73721; AAB32160.1; JOINED.
 CC HSP; P98066; 1TSG.
 CC MGD; MGI:99602; Agc.
 CC InterPro; IPR000436; -
 CC InterPro; IPR000495; -
 CC InterPro; IPR000538; -
 CC InterPro; IPR001304; -
 CC Pfam; PF00193; Xlink; 4.
 CC Pfam; PF00059; lectin_c; 1.
 CC Pfam; PF00084; sush1; 1.
 CC PROSITE; PS00290; IG_MHC; 1.
 CC PROSITE; PS01241; LINK; 4.
 CC PROSITE; PS00615; C-TYPE_LECTIN_1; 1.
 CC PROSITE; PS50041; C-TYPE_LECTIN_2; 1.

Glycoprotein; Cartilage; Proteoglycan; Lectin; Signal; Sushi;
 Repeat; Immunoglobulin domain.
 KW SIGNAL 1 19 POTENTIAL.
 FT CHAIN 20 2132 AGGREGAN CORE PROTEIN.
 FT DOMAIN 44 140 IG-LIKE V-TYPE DOMAIN.
 FT DOMAIN 170 247 LINK 1.
 FT DOMAIN 268 349 LINK 2.
 FT DOMAIN 504 581 LINK 3.
 FT DOMAIN 602 683 LINK 4.
 FT DOMAIN 1918 2044 C-TYPE LECTIN.
 FT DOMAIN 2048 2106 SUSHI.
 FT DOMAIN 48 140 G1-A.
 FT DOMAIN 152 247 G1-B.
 FT DOMAIN 253 349 G1-B'.
 FT DOMAIN 486 580 G2-B.
 FT DOMAIN 587 682 G2-B'.
 FT DOMAIN 685 803 KS.
 FT DOMAIN 805 1231 CS-1.
 FT DOMAIN 1232 1917 CS-2.
 FT DOMAIN 1917 2132 G3.
 FT DISULFID 51 133 BY SIMILARITY.
 FT DISULFID 175 246 BY SIMILARITY.
 FT DISULFID 199 220 BY SIMILARITY.
 FT DISULFID 273 348 BY SIMILARITY.
 FT DISULFID 297 318 BY SIMILARITY.
 FT DISULFID 509 580 BY SIMILARITY.
 FT DISULFID 533 554 BY SIMILARITY.
 FT DISULFID 607 682 BY SIMILARITY.
 FT DISULFID 631 652 BY SIMILARITY.
 FT DISULFID 1922 1933 BY SIMILARITY.
 FT DISULFID 1950 2042 BY SIMILARITY.
 FT DISULFID 2018 2034 BY SIMILARITY.
 FT DISULFID 2049 2092 BY SIMILARITY.
 FT DISULFID 2078 2105 BY SIMILARITY.
 FT CARBOHYD 126 126 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 239 239 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 333 333 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 387 387 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 611 611 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 667 667 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1675 1675 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT SITE 1171 1173 CELL ATTACHMENT SITE (PC64).
 SQ SEQUENCE 2132 AA; 222008 MW; 0B2BCDF6C8DA163 CRC64;
 Query Match 5.0%; Score 140; DB 1; Length 2132;
 Best Local Similarity 20.5%; Pred. No. 2.3;
 Matches 129; Conservative 81; Mismatches 241; Indels 178; Gaps 29;
 QY 4 GNIGNNVNGHILPPAPLPSTQDGAARGGTGHLISSTGALGSRSLFSLRNMSADSVDS 63
 DB 906 GDLSGLPSGGEITETSTSGABEETSLGPSGGDG-LETSTSGVDDVSGIPTGREGLETSSAG 964
 QY 64 -RDIPGLPTNPRLAAATSETCLLGGFEVLHDKGFLDLTLNLTQIGPSAFRVEVQADGTHAA 122
 DB 965 VEDLSGLPSGEGSETSTFS-----GIE-----DISVLPTG-GES-----LETSSAG----- 1004
 QY 123 IGEKNGLEVSTLSPQEWSSLSQSDTECKNRFVFTGGGSGGHPMTVVAS---DIAEART 179
 DB 1005 VGLSGL-----PSGGESLETSSAGAEVDYQLPTERG-----LETSSAGVEDITVLPT 1053
 QY 180 RILAKLPDNGHROPKDVDRSVGVSGASGIDGCVVSETHSTNTSSVRSDPFWVSVG 239
 DB 1054 -----GRE--SLETSSAGVEDVSGLPSC---REGLETSSAGIEDISVFPTAE 1096
 QY 240 ATAAGLAGLAATGIAQALALTPEDDPTTTP-----PSGGDGTETSSAGVEDVSGLPSCGEGLETSSAGVEDLGP 1149
 DB 1097 GLDTSASGGYVSGI-----PSGGDGTETSSAGVEDVSGLPSCGEGLETSSAGVEDLGP 1149
 QY 281 ATKDOLTQEAFAKNPENQKVNKDANGNAIPSGE-----LXDDIVQIAQAQAEAGEV 331
 DB 1150 STRDSL-----ETSAGVDVTG--FPFSGRGPDPETSVSGVGDDDF-----SGLP. 1189

Thu Sep 27 15:29:49 2001

```
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-K12;
RA Aiba H., Baba T., Fujita K., Hayashi K., Honjo A., Horiuchi T.,
RA Ikemoto K., Inada T., Isono K., Itoh S., Itoh T., Kanai K.,
RA Kasai H., Kashimoto K., Kim S., Kimura S., Kitagawa M.,
RA Kitakawa M., Makino K., Masuda S., Miki T., Mizobuchi K., Mori H.,
RA Motomura K., Nakamura Y., Nashimoto H., Nishio Y., Oshima T.,
RA Saito N., Sempel G., Seki Y., Tagami H., Takemoto K., Wada C.,
RA Yamamoto Y., Yano M.;
RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: OUTER MEMBRANE (POTENTIAL).
CC -1- SIMILARITY: STRONG, TO BODETELLA PERTACTIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U36840; AAA79815.1; ALT_SEQ.
DR EMBL; AE000350; AAC75695.1; -.
DR EMBL; D90889; BAA16514.1; ALT_INIT.
DR EMBL; D90890; BAA16518.1; ALT_INIT.
DR EcoGene; EG13213; ypfA.
KW Hypothetical protein; Outer membrane.
SQ SEQUENCE 1569 AA; 162774 MW; 0D006A9A32154596 CRC64;

Query Match 4.98; Score 136; DB 1; Length 1569;
Best Local Similarity 20.98; Pred. No. 2.7;
Matches 127; Conservative 69; Mismatches 259; Indels 154; Gaps 26;

QY 19 APPLSQDTGARGCT--GHLSISGALGSLRSLFPLRNSMADSVDSRDIPGLPTNPSRL 76
DB 579 AKHVQQGGGALIASTTSGTLIEGNSVGDAFY---IRNSEAKNV-----VLENAGSL 628
QY 77 AAATSETCLGGFEVLHDKGPLDILNTOIGP---SAFRVEQVADGTHAAIEKNGLEVSV 133
DB 629 TVVTGSRV---DTIINANGKMDYGVKDVGLNSAGTQTVYASATSKANIKGKGQTVY 685
QY 134 TLSPQE--WSSLQSIDTEKNEFRVFTGGRGGSGHPMTVVASDIAEARTRILAKLPD--- 188
DB 686 GLATEANIESGQIVDGGSTKTHNGGTQVQYNGKAIINTDIVSLQIQIMANGTAEGSI 745
QY 189 -----NHGGRQPKDV-----DTRSVGVGSASGI---DDGVVSETHSTTSSVRS 230
DB 746 INGGQVNVNEGGLAENSVLNDGTLQVRE--KGSATGIQSSQGGALVATRAVTVGTGA 803
QY 231 DKFWVSVGAAGLAGLAATGI--AQALALPEPDDPTTTPDQAAANAESATKDOLTQ 288
DB 804 D-----GVAFSEIQQGAANNILLANGVLAVESD--TSSDKTVQNMGGREIVTKAT- 852
QY 289 EAFKFNQKVNIDANGNAIPGSELXDDIVEQIA-----QQAKEAGEVARQQ--- 335
DB 853 -----ATGTTLTGEE--QIVEGVANETTINDGGIQTVSANGEAIKTKINE 895
QY 336 -----AVESNAQAQRYEDQHARRQEE-----LQLSSGIGYGLSS-----A 371
DB 896 GGTLTVNDNGKATDIVQNSGAALQTSANGIEISGTHOYGTFSISGNLATNMLLENGNL 955
QY 372 LIVAGGIGAGVTTALHRRNPQAGOTTTTHTTVVQQOTGIGIPQHKVALLMPQERRRSDRR 431
DB 956 LVLAGTEARDSTGVKKGAMQNGDSATKVN-----SGG--QYTLGRSKDEFQALARAE 1007
QY 432 DQ--GSVASTHSDSSEVVNVAEYGGGARNLSARHPEEIHVEYADPGYSVTQNF 489
DB 1008 DLQVAGGTAIYAGTLAD-----ASVSGATGSLSLMTPRDNV-----TPVKLE 1050

QY 490 GSGPVTGRLTGPQGIQSTYALL--ANSGLRLMGGLTSG-----CET 532
DB 1051 GAVRTDTSATLTGLNGVDVDTTLADLTAAASRGSVWLNNSNCAGTSNCEYRVNSLLNDGDV 1110
QY 533 AVSSVNAAP 541
DB 1111 YLSAQTAAP 1119
PRT; 2124 AA.

RESULT 7
ID PGCA_RAT STANDARD; PRT; 2124 AA.
AC P07897;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE AGGREGAN CORE PROTEIN PRECURSOR (CARTILAGE-SPECIFIC PROTEOGLYCAN CORE
DE PROTEIN) (CSPCP).
DE AGC1 OR AGC.
GN Rattus norvegicus (Rat).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RN SEQUENCE FROM N.A.
RX MEDLINE=88087070; PubMed=3693370;
RA Doege K., Sasaki M., Horigan E., Hassell J.R., Yamada Y.;
RT "Complete primary structure of the rat cartilage proteoglycan core
RT protein deduced from cDNA clones.";
RL J. Biol. Chem. 262:17757-17767(1987).
RN [2]
RN REVISION TO 698.
RA Doege K., Sasaki M., Horigan E., Hassell J.R., Yamada Y.;
RL J. Biol. Chem. 263:10040-10040(1988).
RN [3]
RN SEQUENCE OF 1856-2124 FROM N.A.
RX MEDLINE=86250698; PubMed=2424893;
RA Doege K., Fernandez P., Hassell J.R., Sasaki M., Yamada Y.;
RT "Partial cDNA sequence encoding a globular domain at the C terminus
RT of the rat cartilage proteoglycan.";
RL J. Biol. Chem. 261:8108-8111(1986).
CC -1- FUNCTION: THIS PROTEOGLYCAN IS A MAJOR COMPONENT OF EXTRACELLULAR
CC MATRIX OF CARTILAGINOUS TISSUES. A MAJOR FUNCTION OF THIS PROTEIN
CC IS TO RESIST COMPRESSION IN CARTILAGE. IT BINDS AVIDLY TO
CC HYALURONIC ACID VIA AN AMINO-TERMINAL GLOBULAR REGION. MAY PLAY A
CC REGULATORY ROLE IN THE MATRIX ASSEMBLY OF THE CARTILAGE.
CC -1- SUBCELLULAR LOCATION: SECRETED; EXTRACELLULAR MATRIX (BY
CC SIMILARITY).
CC -1- DOMAIN: TWO GLOBULAR DOMAINS, G1 AND G2, COMPRISE THE AMINO
CC TERMINUS OF THE PROTEOGLYCAN, WHILE ANOTHER GLOBULAR REGION, G3,
CC MAKES UP THE COOH TERMINUS. G1 CONTAINS LINK DOMAINS AND THUS
CC CONSISTS OF THREE DISULFIDE-BONDED LOOP STRUCTURES DESIGNATED AS
CC THE A, B, B' MOTIFS. G2 IS SIMILAR TO G1. THE KERATAN SULFATE (KS)
CC AND THE CHONDROITIN SULFATE (CS) ATTACHMENT DOMAINS LIE BETWEEN G2
CC AND G3.
CC -1- PTM: CONTAINS MOSTLY CHONDROITIN SULFATE, BUT ALSO KERATAN SULFATE
CC CHAINS. N-LINKED AND O-LINKED OLIGOSACCHARIDES.
CC -1- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE V-TYPE DOMAIN.
CC -1- SIMILARITY: CONTAINS 4 LINK DOMAINS.
CC -1- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 SUSHI (SCR) REPEAT.
CC -1- SIMILARITY: BELONGS TO THE AGGREGAN/VERSICAN PROTEOGLYCAN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M13518; AAA41836.1; -.
```

```

DR EMBL; J03485; AAA21000.1; ALT_SEQ.
DR PIR; A23835; A23835.
DR PIR; A28452; A28452.
DR HSP; P98066; ITS6.
DR InterPro; IPR000436; -.
DR InterPro; IPR000495; -.
DR InterPro; IPR000538; -.
DR InterPro; IPR001304; -.
DR Pfam; PF00193; Xlink; 4.
DR Pfam; PF00059; lectin_c; 1.
DR Pfam; PF00084; sushi; 1.
DR PROSITE; PS00290; IG_MHC; 1.
DR PROSITE; PS01241; LINK; 4.
DR PROSITE; PS00615; C-TYPE_LECTIN_1; 1.
DR PROSITE; PS50041; C-TYPE_LECTIN_2; 1.
KW Glycoprotein; Cartilage; Proteoglycan; Lectin; Signal; Sushi;
KW Repeat; Immunoglobulin domain.
FT SIGNAL 1 19 POTENTIAL.
FT CHAIN 20 2124 AGGRECAN CORE PROTEIN.
FT DOMAIN 44 140 IG-LIKE V-TYPE DOMAIN.
FT DOMAIN 170 247 LINK 1.
FT DOMAIN 268 349 LINK 2.
FT DOMAIN 504 581 LINK 3.
FT DOMAIN 602 683 LINK 4.
FT DOMAIN 1910 2036 C-TYPE LECTIN.
FT DOMAIN 2040 2098 SUSHI.
FT DOMAIN 48 140 G1-A.
FT DOMAIN 152 247 G1-B.
FT DOMAIN 253 349 G1-B'.
FT DOMAIN 486 580 G2-B.
FT DOMAIN 587 682 G2-B'.
FT DOMAIN 685 798 KS.
FT DOMAIN 801 1226 CS-1.
FT DOMAIN 1227 1909 CS-2.
FT DOMAIN 1910 2124 CS-3.
FT DISULFID 51 133 BY SIMILARITY.
FT DISULFID 175 246 BY SIMILARITY.
FT DISULFID 199 220 BY SIMILARITY.
FT DISULFID 273 348 BY SIMILARITY.
FT DISULFID 297 318 BY SIMILARITY.
FT DISULFID 509 580 BY SIMILARITY.
FT DISULFID 533 554 BY SIMILARITY.
FT DISULFID 607 682 BY SIMILARITY.
FT DISULFID 631 652 BY SIMILARITY.
FT DISULFID 1914 1925 BY SIMILARITY.
FT DISULFID 1942 2034 BY SIMILARITY.
FT DISULFID 2010 2026 BY SIMILARITY.
FT DISULFID 2041 2084 BY SIMILARITY.
FT DISULFID 2070 2097 BY SIMILARITY.
FT CARBOHYD 126 126 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 239 239 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 333 333 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 387 387 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 611 611 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 667 667 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1842 1842 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 2124 AA; 221117 MW; E30BBE61593A34B1 CRC64;

Query Match 4.8%; Score 135.5; DB 1; Length 2124;
Best Local Similarity 20.7%; Pred. No. 4.2;
Matches 138; Conservative 77; Mismatches 204; Indels 247; Gaps 33;

QY 17 PPAPPLPSQTD----GAARGGTGHL-----ISTGALGRSLFSPLRNMSVDSR 64
DB 880 PTVDRLPSGSGESLEGSASAGTGLSGLPSSGEITTSASG-----TE 922
QY 65 DIPGLPTNPSRLAAATSETCILGGFEVL-HDKGPDILNTQIGPSAFRVEQADGTHAI 123
DB 923 EISGLPSGGDDLETSTSG---IDGASVLTPTGRGGL-----ETSASGVEDLS 965
QY 124 GKNGLEVSVTLSPOBWSLSQSDT--EGKNRFVFTGGRG-----GSGHPMVTVASDIA 175

```

```

Db 966 GLPSGEGSETST----SGIEDISVLPTGESPETASGVGLSLPSGSGESLETASGVE 1021
QY 176 EARTRILAKLPDNGHGRQPKVDYRFSVGVGSAGIDDDGVVSET---HTSTNSSVRS DP 232
Db 1022 DV-TQL-----PTERGGLT-----SASGIEDITVLPTGRENLETASGVE--- 1061
QY 233 KFWVSVAIAAGLAGL--AATGI-----AQALALT-----PEPDDPTTDP- 271
Db 1062 ----DVSGLPSCKEGLETASGIEDISVPTAEAGLETASGGYVSGIPSGEDGTST 1117
QY 272 -----DQANAAS-----ATKDQLTOFAKNPENQKNVIDANGNAIPSGE- 312
Db 1118 GVEGVSLPSGCGELETASGVVEDLGLPTRDSL-----ETSASGVDTG--YPSGRE 1167
QY 313 -----LXDDIVEIOAQOAKEAGEVARQOAVESNAQOQRYEDOHARQOELQLSSG- 363
Db 1168 DTETSPVGGDDL-----SGLPSGQEGLETASGAEDLGGLPSCGKEDLVGSASGA 1217
QY 364 -----IGYLSSALIVAGGICAGVTTALHRR 389
Db 1218 LDFGKLPSGTLSGGQTPEASGLPSGFSGEYSVDIGSGFSSGLPDFSGLPSGPTVSLVD 1277
QY 390 NQPAEQTTTTTHVVQQOT-----GGIPQHKVAMPQERRRFSRRRDSQGSVAST 440
Db 1278 STLVEVITATTASELEGRGTISVSGSGESGPPLSL-----DSSADISGLPSGT 1327
QY 441 HNSDSSEVVNPNYAEVGGGARNLSAHQPEHIYDEVAADPGYVIONFSGSGPVTCRLIC 500
Db 1328 ELSGQTSGSLDVSGETSG-----FFDVSGQPFSG-----SGTGGTSGIPE 1368
QY 501 TPGQGIQS--TVALLANS--GLRLGMG---GLTS--GGETA--VSSVNA 540
Db 1369 VSGQAVRSPDTTEISELSGLSSGQPDVSGEGSGLIFSGSGSGITSVSGTSGISDL 1428
QY 541 PTPGPV 546
Db 1429 PSGFPV 1434

RESULT 8
APMO_PIG STANDARD; PRT; 1150 AA.
AC P12021;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE APOMUCIN (MUCIN CORE PROTEIN) (FRAGMENT).
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-Submaxillary gland;
RA MEDLINE=91236743; PubMed=2033060;
RX Eckhardt A.E., Timpte C.S., Abernethy J.L., Zhao Y., Hill R.L.;
RT "Porcine submaxillary mucin contains a cysteine-rich,
carboxyl-terminal domain in addition to a highly repetitive,
glycosylated domain.";
RL J. Biol. Chem. 266:9678-9686(1991).
RN [2]
RP SEQUENCE OF 1-503 FROM N.A.
RC TISSUE-Submaxillary gland;
RX MEDLINE=88087170; PubMed=2826455;
RA Timpte C.S., Eckhardt A.E., Abernethy J.L., Hill R.L.;
RT "Porcine submaxillary gland apomucin contains tandemly repeated,
identical sequences of 81 residues.";
RL J. Biol. Chem. 263:1081-1088(1988).
RN [3]
RP SEQUENCE OF 45-80.
RC TISSUE-Submaxillary gland;
RX MEDLINE=87280230; PubMed=3611111;
RA Eckhardt A.E., Timpte C.S., Abernethy J.L., Toumadje A.,

```



```

Schizophyllaceae; Schizophyllum.
NCBI_TaxID=5334;
[1]
RP SEQUENCE FROM N.A.
RC STRAIN=UVM 4-40;
RX MEDLINE=92357793; PubMed=1353886;
RA Stankis M.M., Specht C.A., Yang H., Giasson L., Ullrich R.C.,
RA Novotny C.P.;
RT "The A alpha mating locus of Schizophyllum commune encodes two
RT dissimilar multiallelic homeodomain proteins.";
RL Proc. Natl. Acad. Sci. U.S.A. 89:7169-7173(1992).
CC -1- FUNCTION: SPECIES A-ALPHA-4 MATING-TYPE. MAY REGULATE THE
CC EXPRESSION OF GENES SPECIFIC TO THE HOMOKARYOTIC CELL TYPE.
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -1- DEVELOPMENTAL STAGE: EXPRESSED CONSTITUTIVELY IN HOMOKARYONS.
CC -1- SIMILARITY: BELONGS TO THE TALE/M-ATYP FAMILY OF HOMEOBOX
CC PROTEINS.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
DR PIR: D37271; D37271.
DR InterPro: IPR001356;
DR PROSITE: PS00027; HOMEOBOX_1; FALSE-NEG.
DR PROSITE: PS00071; HOMEOBOX_2; FALSE-NEG.
KW Homeobox; DNA-binding; Transcription regulation; Nuclear protein.
FT DNA_BIND 110 182 HOMEOBOX (TALE-TYPE).
FT DOMAIN 370 402 ASP/GLU-RICH (ACIDIC).
FT DOMAIN 436 469 ASP/GLU-RICH (ACIDIC).
SQ SEQUENCE 940 AA; 101856 MW; 4B99CBAEDB39621E CRC64;
-----
Query Match
Best Local Similarity 4.7%; Score 130.5; DB 1; Length 940;
Matches 113; Conservative 63; Mismatches 205; Indels 121; Gaps 21;
QY 23 PSQTGARGGTHGISLSTGALGSLRSLSP-----LRNSMADSVDSR--DIPGLPTN 72
D 470 PVKIAGAKRGNDDEVSPLAKKPR-IFSPVPRPQAIRVSLSPAPSSRGSTPTSPVS 528
QY 73 PS---RLAATSEFCLLGGFV-LHDKGLDILNTQGPSAFRVEVQADGTH-AAIGEK 126
D 529 PSPKAKRPAQATS---LLASHPMKKREKLQELRKAGLAPPSAPVLMGPDGVPGLGTVRSR 585
QY 127 NGLEVSVTLSPOEWSLQSIDTEGKNRFVFTGGRGSGHPMVTVASDAEARTILAKLD 186
D 586 SSPSVS---SPSPSVSLPLPSRG---VPSGGIKVTGPTFWVNDL-EAHT----- 630
QY 187 PDNHGGRQPKD-----VD-----TRSVGVGSAGIDDDGVVSETHST 223
D 631 -----QAPRLTAATKSSAGSVDAVPLPGKRSRLTRSPSSIS-----SACST 675
QY 224 TNSVSRDPKFWVSGAIAAGLAGLAATGIAQALALTPEPDPTTDDQAAANAESATK 283
D 676 SSSGSDTDSLSFVT-----SDATDIT-EPDEATTADETTTQSTASSSR 718
QY 284 DLTQEAFAKNPNQKVNID-----ANGNAIPSGEL-----XDDIVEQITA 322
D 719 DTTSQKRMPPSIDPRDPALMSYDLSPPADGRHLHPSDGLRPSAFVPTKLDVRVANLA 778
QY 323 QQAKEAGEVARQQAQOQRYEDQHARQEELQLSSGIGYGLSALIIVAG---GIG 379
D 779 QNPARHWSAKSPFASHAAAPVSYHHATGSIASPAQVAFEGQLTSVLATGQKAGNA 838
QY 380 AGVTTALHRRNOPASQTTTTTHTVVQQQTGGIPO-----HKVALMPQRRRRFSRDSQ 434
D 839 RRRRTFVQRRVTPKAQETSEPSLSVDGILSSGLADVCREAPKAPKAPKNDRRYLERRR 898
-----
QY 435 GSVASTHMSDSSEVVNPAEV 456
| : | | : | | :
DB 899 LSKSSP--VDSADTVTRLAEI 918
| : | | : | | :
RESULT 13
ALSL_CANAL STANDARD; PRT; 1260 AA.
ID P46590;
AC 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE AGGLUTININ-LIKE PROTEIN 1 PRECURSOR.
GN ALSI.
OS Candida albicans (Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX NCBI_TaxID=5476;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 11651 / B792;
RX MEDLINE=95273392; PubMed=7752895;
RA Hoyer L.L., Scherer S., Shatzman A.R., Livi G.P.;
RA "Candida albicans ALSI: domains related to a Saccharomyces cerevisiae
RT sexual agglutinin separated by a repeating motif.";
RL Mol. Microbiol. 15:39-54(1995).
CC -1- FUNCTION: MAY PLAY A ROLE IN ADHESION AND PATHOGENESIS.
CC -1- PTM: N-GLYCOSYLATED AND O-GLYCOSYLATED (POTENTIAL).
CC -1- SIMILARITY: TO YEAST SAGI.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL: L25902; AAC41649.2; -.
KW Cell adhesion; Glycoprotein; Repeat; Signal.
FT SIGNAL 1 17
FT CHAIN 18 1260 AGGLUTININ-LIKE PROTEIN 1.
FT DOMAIN 433 792 10 X 36 AA TANDEM REPEATS.
FT REPEAT 433 468 1-1.
FT REPEAT 469 504 1-2.
FT REPEAT 505 540 1-3.
FT REPEAT 541 576 1-4.
FT REPEAT 577 612 1-5.
FT REPEAT 613 648 1-6.
FT REPEAT 649 684 1-7.
FT REPEAT 685 720 1-8.
FT REPEAT 721 756 1-9.
FT REPEAT 757 792 1-10.
FT DOMAIN 983 1152 2 X 26 AA APPROXIMATE REPEATS.
FT REPEAT 983 1043 2-1.
FT REPEAT 1092 1152 2-2.
FT DOMAIN 399 404 POLY-THR.
FT DOMAIN 408 418 POLY-THR.
FT DOMAIN 450 455 POLY-THR.
FT DOMAIN 486 491 POLY-THR.
FT DOMAIN 522 527 POLY-THR.
FT DOMAIN 558 563 POLY-THR.
FT DOMAIN 594 599 POLY-THR.
FT DOMAIN 630 635 POLY-THR.
FT DOMAIN 666 671 POLY-THR.
FT DOMAIN 702 707 POLY-THR.
FT DOMAIN 738 743 POLY-THR.
FT DOMAIN 774 779 POLY-THR.
FT DOMAIN 874 877 POLY-SER.
FT CARBOHYD 471 471 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 579 579 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 615 615 N-LINKED (GLCNAC. . .) (POTENTIAL).
```


OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 ON NCBI_TaxID=7227;
 RX SEQUENCE FROM N.A.
 RP MEDLINE=86079539; PubMed=3935325;
 RA Wharton K.A., Johansen K.M., Xu T., Artavanis-Tsakonas S.;
 RT "Nucleotide sequence from the neurogenic locus notch implies a gene
 RT product that shares homology with proteins containing EGF-like
 RT repeats.";
 RL Cell 43:567-581(1985).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=OREGON-R;
 RX MEDLINE=87064624; PubMed=3097517;
 RA Kidd S., Kelley M.R., Young M.W.;
 RT "Sequence of the notch locus of Drosophila melanogaster: relationship
 RT of the encoded protein to mammalian clotting and growth factors.";
 RL Mol. Cell. Biol. 6:3094-3108(1986).
 RN [3]
 RP SEQUENCE OF 2505-2611 FROM N.A.
 RX MEDLINE=85093929; PubMed=2981631;
 RA Wharton K.A., Yedvobnick B., Finnerty V.G., Artavanis-Tsakonas S.;
 RT "Opa: a novel family of transcribed repeats shared by the Notch locus
 RT and other developmentally regulated loci in D. melanogaster.";
 RL Cell 40:55-62(1985).
 RN [4]
 RP SEQUENCE OF 1-8 FROM N.A.
 RX MEDLINE=87257846; PubMed=3037327;
 RA Kelley M.R., Kidd S., Berg R.L., Young M.W.;
 RT "Restriction of P-element insertions at the Notch locus of Drosophila
 RT melanogaster.";
 RL Mol. Cell. Biol. 7:1545-1548(1987).
 RN [5]
 RP REVIEW.
 RA Harris W.A.;
 RT "Many cell types specified by Notch function.";
 RL Curr. Biol. 1:120-122(1991).
 CC -!- FUNCTION: NOTCH PROTEIN IS ESSENTIAL FOR PROPER DIFFERENTIATION OF
 CC ECTODERM.
 CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
 CC -!- MISCELLANEOUS: SEPARATION OF NEUROBLASTS FROM THE ECTODERM INTO
 CC THE INNER PART OF EMBRYO IS ONE OF THE FIRST STEPS OF CNS
 CC DEVELOPMENT IN INSECTS, THIS PROCESS IS UNDER CONTROL OF THE
 CC NEUROGENIC GENES.
 CC -!- SIMILARITY: HIGH, WITH OTHER NOTCH-TYPE PROTEINS.
 CC -!- SIMILARITY: CONTAINS 36 EGF-LIKE DOMAINS.
 CC -!- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.
 CC -!- SIMILARITY: CONTAINS 6 ANK REPEATS.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; M16152; AAB59220.1; -
 CC EMBL; M16153; AAB59220.1; JOINED.
 CC EMBL; M16149; AAB59220.1; JOINED.
 CC EMBL; M16150; AAB59220.1; JOINED.
 CC EMBL; M16151; AAB59220.1; JOINED.
 CC EMBL; K03508; AAA28725.1; -
 CC EMBL; M13689; AAA28725.1; JOINED.
 CC EMBL; K03507; AAA28725.1; JOINED.
 CC EMBL; M12175; AAA74496.1; -
 CC EMBL; M16025; AAA28726.1; -
 CC PIR; A24420; A24420.
 CC PIR; A24768; A24768.
 CC PIR; A05267; A05267.

DR HSSP; P00740; LIXA.
 DR FlyBase; FBgn0004647; N.
 DR InterPro; IPR000152; -
 DR InterPro; IPR000561; -
 DR InterPro; IPR000800; -
 DR InterPro; IPR001438; -
 DR InterPro; IPR001881; -
 DR InterPro; IPR002110; -
 DR Pfam; PF00008; EGF; 36.
 DR Pfam; PF00023; ank; 6.
 DR Pfam; PF00066; notch; 3.
 DR PRINTS; PR00010; EGFBL00D.
 DR PROSITE; PS50088; ANK_REPEAT; 5.
 DR PROSITE; PS50297; ANK_REPEAT_REGION; 1.
 DR PROSITE; PS00010; ASK_HYDROXYL; 22.
 DR PROSITE; PS00022; EGF_1; 34.
 DR PROSITE; PS01186; EGF_2; 28.
 DR PROSITE; PS01187; EGF_CA; 22.
 KW Differentiation; Neurogenesis; Repeat; ANK repeat; EGF-like domain;
 KW Transmembrane; Signal; Glycoprotein.
 FT SIGNAL 1 44
 FT CHAIN 45 2703
 FT DOMAIN 45 1745
 FT TRANSMEM 1746 1766
 FT DOMAIN 1767 2703
 FT DOMAIN 58 95
 FT DOMAIN 96 136
 FT DOMAIN 139 176
 FT DOMAIN 177 215
 FT DOMAIN 217 253
 FT DOMAIN 255 291
 FT DOMAIN 293 329
 FT DOMAIN 331 370
 FT DOMAIN 372 408
 FT DOMAIN 409 447
 FT DOMAIN 449 486
 FT DOMAIN 488 524
 FT DOMAIN 526 562
 FT DOMAIN 564 600
 FT DOMAIN 602 637
 FT DOMAIN 639 675
 FT DOMAIN 677 713
 FT DOMAIN 715 751
 FT DOMAIN 753 789
 FT DOMAIN 791 827
 FT DOMAIN 829 865
 FT DOMAIN 867 905
 FT DOMAIN 907 944
 FT DOMAIN 946 982
 FT DOMAIN 984 1020
 FT DOMAIN 1022 1058
 FT DOMAIN 1060 1096
 FT DOMAIN 1098 1134
 FT DOMAIN 1136 1181
 FT DOMAIN 1183 1219
 FT DOMAIN 1221 1257
 FT DOMAIN 1259 1295
 FT DOMAIN 1297 1335
 FT DOMAIN 1337 1373
 FT DOMAIN 1375 1412
 FT DOMAIN 1415 1451
 FT REPEAT 1475 1513
 FT REPEAT 1514 1553
 FT REPEAT 1554 1593
 FT REPEAT 1591 1627
 FT REPEAT 1629 1665
 FT REPEAT 1667 1703
 FT REPEAT 1705 1741
 FT REPEAT 1743 1779
 FT REPEAT 1781 1817
 FT REPEAT 1819 1855
 FT REPEAT 1857 1893
 FT REPEAT 1895 1931
 FT REPEAT 1933 1969
 FT REPEAT 1971 2007
 FT REPEAT 2009 2045
 FT REPEAT 2047 2083
 FT REPEAT 2085 2121
 FT DOMAIN 2123 2568
 FT DISULFID 62 73
 FT DISULFID 75 83
 FT DISULFID 86 94
 FT DISULFID 97 105
 FT DISULFID 106 114
 FT DISULFID 115 123
 FT DISULFID 124 132
 FT DISULFID 133 141
 FT DISULFID 140 148
 FT DISULFID 149 157
 FT DISULFID 156 164
 FT DISULFID 165 173
 FT DISULFID 172 180
 FT DISULFID 180 188
 FT DISULFID 188 196
 FT DISULFID 196 204
 FT DISULFID 204 212
 FT DISULFID 212 220
 FT DISULFID 220 228
 FT DISULFID 228 236
 FT DISULFID 236 244
 FT DISULFID 244 252
 FT DISULFID 252 260
 FT DISULFID 260 268
 FT DISULFID 268 276
 FT DISULFID 276 284
 FT DISULFID 284 292
 FT DISULFID 292 300
 FT DISULFID 300 308
 FT DISULFID 308 316
 FT DISULFID 316 324
 FT DISULFID 324 332
 FT DISULFID 332 340
 FT DISULFID 340 348
 FT DISULFID 348 356
 FT DISULFID 356 364
 FT DISULFID 364 372
 FT DISULFID 372 380
 FT DISULFID 380 388
 FT DISULFID 388 396
 FT DISULFID 396 404
 FT DISULFID 404 412
 FT DISULFID 412 420
 FT DISULFID 420 428
 FT DISULFID 428 436
 FT DISULFID 436 444
 FT DISULFID 444 452
 FT DISULFID 452 460
 FT DISULFID 460 468
 FT DISULFID 468 476
 FT DISULFID 476 484
 FT DISULFID 484 492
 FT DISULFID 492 500
 FT DISULFID 500 508
 FT DISULFID 508 516
 FT DISULFID 516 524
 FT DISULFID 524 532
 FT DISULFID 532 540
 FT DISULFID 540 548
 FT DISULFID 548 556
 FT DISULFID 556 564
 FT DISULFID 564 572
 FT DISULFID 572 580
 FT DISULFID 580 588
 FT DISULFID 588 596
 FT DISULFID 596 604
 FT DISULFID 604 612
 FT DISULFID 612 620
 FT DISULFID 620 628
 FT DISULFID 628 636
 FT DISULFID 636 644
 FT DISULFID 644 652
 FT DISULFID 652 660
 FT DISULFID 660 668
 FT DISULFID 668 676
 FT DISULFID 676 684
 FT DISULFID 684 692
 FT DISULFID 692 700
 FT DISULFID 700 708
 FT DISULFID 708 716
 FT DISULFID 716 724
 FT DISULFID 724 732
 FT DISULFID 732 740
 FT DISULFID 740 748
 FT DISULFID 748 756
 FT DISULFID 756 764
 FT DISULFID 764 772
 FT DISULFID 772 780
 FT DISULFID 780 788
 FT DISULFID 788 796
 FT DISULFID 796 804
 FT DISULFID 804 812
 FT DISULFID 812 820
 FT DISULFID 820 828
 FT DISULFID 828 836
 FT DISULFID 836 844
 FT DISULFID 844 852
 FT DISULFID 852 860
 FT DISULFID 860 868
 FT DISULFID 868 876
 FT DISULFID 876 884
 FT DISULFID 884 892
 FT DISULFID 892 900
 FT DISULFID 900 908
 FT DISULFID 908 916
 FT DISULFID 916 924
 FT DISULFID 924 932
 FT DISULFID 932 940
 FT DISULFID 940 948
 FT DISULFID 948 956
 FT DISULFID 956 964
 FT DISULFID 964 972
 FT DISULFID 972 980
 FT DISULFID 980 988
 FT DISULFID 988 996
 FT DISULFID 996 1004
 FT DISULFID 1004 1012
 FT DISULFID 1012 1020
 FT DISULFID 1020 1028
 FT DISULFID 1028 1036
 FT DISULFID 1036 1044
 FT DISULFID 1044 1052
 FT DISULFID 1052 1060
 FT DISULFID 1060 1068
 FT DISULFID 1068 1076
 FT DISULFID 1076 1084
 FT DISULFID 1084 1092
 FT DISULFID 1092 1100
 FT DISULFID 1100 1108
 FT DISULFID 1108 1116
 FT DISULFID 1116 1124
 FT DISULFID 1124 1132
 FT DISULFID 1132 1140
 FT DISULFID 1140 1148
 FT DISULFID 1148 1156
 FT DISULFID 1156 1164
 FT DISULFID 1164 1172
 FT DISULFID 1172 1180
 FT DISULFID 1180 1188
 FT DISULFID 1188 1196
 FT DISULFID 1196 1204
 FT DISULFID 1204 1212
 FT DISULFID 1212 1220
 FT DISULFID 1220 1228
 FT DISULFID 1228 1236
 FT DISULFID 1236 1244
 FT DISULFID 1244 1252
 FT DISULFID 1252 1260
 FT DISULFID 1260 1268
 FT DISULFID 1268 1276
 FT DISULFID 1276 1284
 FT DISULFID 1284 1292
 FT DISULFID 1292 1300
 FT DISULFID 1300 1308
 FT DISULFID 1308 1316
 FT DISULFID 1316 1324
 FT DISULFID 1324 1332
 FT DISULFID 1332 1340
 FT DISULFID 1340 1348
 FT DISULFID 1348 1356
 FT DISULFID 1356 1364
 FT DISULFID 1364 1372
 FT DISULFID 1372 1380
 FT DISULFID 1380 1388
 FT DISULFID 1388 1396
 FT DISULFID 1396 1404
 FT DISULFID 1404 1412
 FT DISULFID 1412 1420
 FT DISULFID 1420 1428
 FT DISULFID 1428 1436
 FT DISULFID 1436 1444
 FT DISULFID 1444 1452
 FT DISULFID 1452 1460
 FT DISULFID 1460 1468
 FT DISULFID 1468 1476
 FT DISULFID 1476 1484
 FT DISULFID 1484 1492
 FT DISULFID 1492 1500
 FT DISULFID 1500 1508
 FT DISULFID 1508 1516
 FT DISULFID 1516 1524
 FT DISULFID 1524 1532
 FT DISULFID 1532 1540
 FT DISULFID 1540 1548
 FT DISULFID 1548 1556
 FT DISULFID 1556 1564
 FT DISULFID 1564 1572
 FT DISULFID 1572 1580
 FT DISULFID 1580 1588
 FT DISULFID 1588 1596
 FT DISULFID 1596 1604
 FT DISULFID 1604 1612
 FT DISULFID 1612 1620
 FT DISULFID 1620 1628
 FT DISULFID 1628 1636
 FT DISULFID 1636 1644
 FT DISULFID 1644 1652
 FT DISULFID 1652 1660
 FT DISULFID 1660 1668
 FT DISULFID 1668 1676
 FT DISULFID 1676 1684
 FT DISULFID 1684 1692
 FT DISULFID 1692 1700
 FT DISULFID 1700 1708
 FT DISULFID 1708 1716
 FT DISULFID 1716 1724
 FT DISULFID 1724 1732
 FT DISULFID 1732 1740
 FT DISULFID 1740 1748
 FT DISULFID 1748 1756
 FT DISULFID 1756 1764
 FT DISULFID 1764 1772
 FT DISULFID 1772 1780
 FT DISULFID 1780 1788
 FT DISULFID 1788 1796
 FT DISULFID 1796 1804
 FT DISULFID 1804 1812
 FT DISULFID 1812 1820
 FT DISULFID 1820 1828
 FT DISULFID 1828 1836
 FT DISULFID 1836 1844
 FT DISULFID 1844 1852
 FT DISULFID 1852 1860
 FT DISULFID 1860 1868
 FT DISULFID 1868 1876
 FT DISULFID 1876 1884
 FT DISULFID 1884 1892
 FT DISULFID 1892 1900
 FT DISULFID 1900 1908
 FT DISULFID 1908 1916
 FT DISULFID 1916 1924
 FT DISULFID 1924 1932
 FT DISULFID 1932 1940
 FT DISULFID 1940 1948
 FT DISULFID 1948 1956
 FT DISULFID 1956 1964
 FT DISULFID 1964 1972
 FT DISULFID 1972 1980
 FT DISULFID 1980 1988
 FT DISULFID 1988 1996
 FT DISULFID 1996 2004
 FT DISULFID 2004 2012
 FT DISULFID 2012 2020
 FT DISULFID 2020 2028
 FT DISULFID 2028 2036
 FT DISULFID 2036 2044
 FT DISULFID 2044 2052
 FT DISULFID 2052 2060
 FT DISULFID 2060 2068
 FT DISULFID 2068 2076
 FT DISULFID 2076 2084
 FT DISULFID 2084 2092
 FT DISULFID 2092 2100
 FT DISULFID 2100 2108
 FT DISULFID 2108 2116
 FT DISULFID 2116 2124
 FT DISULFID 2124 2132
 FT DISULFID 2132 2140
 FT DISULFID 2140 2148
 FT DISULFID 2148 2156
 FT DISULFID 2156 2164
 FT DISULFID 2164 2172
 FT DISULFID 2172 2180
 FT DISULFID 2180 2188
 FT DISULFID 2188 2196
 FT DISULFID 2196 2204
 FT DISULFID 2204 2212
 FT DISULFID 2212 2220
 FT DISULFID 2220 2228
 FT DISULFID 2228 2236
 FT DISULFID 2236 2244
 FT DISULFID 2244 2252
 FT DISULFID 2252 2260
 FT DISULFID 2260 2268
 FT DISULFID 2268 2276
 FT DISULFID 2276 2284
 FT DISULFID 2284 2292
 FT DISULFID 2292 2300
 FT DISULFID 2300 2308
 FT DISULFID 2308 2316
 FT DISULFID 2316 2324
 FT DISULFID 2324 2332
 FT DISULFID 2332 2340
 FT DISULFID 2340 2348
 FT DISULFID 2348 2356
 FT DISULFID 2356 2364
 FT DISULFID 2364 2372
 FT DISULFID 2372 2380
 FT DISULFID 2380 2388
 FT DISULFID 2388 2396
 FT DISULFID 2396 2404
 FT DISULFID 2404 2412
 FT DISULFID 2412 2420
 FT DISULFID 2420 2428
 FT DISULFID 2428 2436
 FT DISULFID 2436 2444
 FT DISULFID 2444 2452
 FT DISULFID 2452 2460
 FT DISULFID 2460 2468
 FT DISULFID 2468 2476
 FT DISULFID 2476 2484
 FT DISULFID 2484 2492
 FT DISULFID 2492 2500
 FT DISULFID 2500 2508
 FT DISULFID 2508 2516
 FT DISULFID 2516 2524
 FT DISULFID 2524 2532
 FT DISULFID 2532 2540
 FT DISULFID 2540 2548
 FT DISULFID 2548 2556
 FT DISULFID 2556 2564
 FT DISULFID 2564 2572
 FT DISULFID 2572 2580
 FT DISULFID 2580 2588
 FT DISULFID 2588 2596
 FT DISULFID 2596 2604
 FT DISULFID 2604 2612
 FT DISULFID 2612 2620
 FT DISULFID 2620 2628
 FT DISULFID 2628 2636
 FT DISULFID 2636 2644
 FT DISULFID 2644 2652
 FT DISULFID 2652 2660
 FT DISULFID 2660 2668
 FT DISULFID 2668 2676
 FT DISULFID 2676 2684
 FT DISULFID 2684 2692
 FT DISULFID 2692 2700
 FT DISULFID 2700 2708
 FT DISULFID 2708 2716
 FT DISULFID 2716 2724
 FT DISULFID 2724 2732
 FT DISULFID 2732 2740
 FT DISULFID 2740 2748
 FT DISULFID 2748 2756
 FT DISULFID 2756 2764
 FT DISULFID 2764 2772
 FT DISULFID 2772 2780
 FT DISULFID 2780 2788
 FT DISULFID 2788 2796
 FT DISULFID 2796 2804
 FT DISULFID 2804 2812
 FT DISULFID 2812 2820
 FT DISULFID 2820 2828
 FT DISULFID 2828 2836
 FT DISULFID 2836 2844
 FT DISULFID 2844 2852
 FT DISULFID 2852 2860
 FT DISULFID 2860 2868
 FT DISULFID 2868 2876
 FT DISULFID 2876 2884
 FT DISULFID 2884 2892
 FT DISULFID 2892 2900
 FT DISULFID 2900 2908
 FT DISULFID 2908 2916
 FT DISULFID 2916 2924
 FT DISULFID 2924 2932
 FT DISULFID 2932 2940
 FT DISULFID 2940 2948
 FT DISULFID 2948 2956
 FT DISULFID 2956 2964
 FT DISULFID 2964 2972
 FT DISULFID 2972 2980
 FT DISULFID 2980 2988
 FT DISULFID 2988 2996
 FT DISULFID 2996 3004
 FT DISULFID 3004 3012
 FT DISULFID 3012 3020
 FT DISULFID 3020 3028
 FT DISULFID 3028 3036
 FT DISULFID 3036 3044
 FT DISULFID 3044 3052
 FT DISULFID 3052 3060
 FT DISULFID 3060 3068
 FT DISULFID 3068 3076
 FT DISULFID 3076 3084
 FT DISULFID 3084 3092
 FT DISULFID 3092 3100
 FT DISULFID 3100 3108
 FT DISULFID 3108 3116
 FT DISULFID 3116 3124
 FT DISULFID 3124 3132
 FT DISULFID 3132 3140
 FT DISULFID 3140 3148
 FT DISULFID 3148 3156
 FT DISULFID 3156 3164
 FT DISULFID 3164 3172
 FT DISULFID 3172 3180
 FT DISULFID 3180 3188
 FT DISULFID 3188 3196
 FT DISULFID 3196 3204
 FT DISULFID 3204 3212
 FT DISULFID 3212 3220
 FT DISULFID 3220 3228
 FT DISULFID 3228 3236
 FT DISULFID 3236 3244
 FT DISULFID 3244 3252
 FT DISULFID 3252 3260
 FT DISULFID 3260 3268
 FT DISULFID 3268 3276
 FT DISULFID 3276 3284
 FT DISULFID 3284 3292
 FT DISULFID 3292 3300
 FT DISULFID 3300 3308
 FT DISULFID 3308 3316
 FT DISULFID 3316 3324
 FT DISULFID 3324 3332
 FT DISULFID 3332 3340
 FT DISULFID 3340 3348
 FT DISULFID 3348 3356
 FT DISULFID 3356 3364
 FT DISULFID 3364 3372
 FT DISULFID 3372 3380
 FT DISULFID 3380 3388
 FT DISULFID 3388 3396
 FT DISULFID 3396 3404
 FT DISULFID 3404 3412
 FT DISULFID 3412 3420
 FT DISULFID 3420 3428
 FT DISULFID 3428 3436
 FT DISULFID 3436 3444
 FT DISULFID 3444 3452
 FT DISULFID 3452 3460
 FT DISULFID 3460 3468
 FT DISULFID 3468 3476
 FT DISULFID 3476 3484
 FT DISULFID 3484 3492
 FT DISULFID 3492 3500
 FT DISULFID 3500 3508
 FT DISULFID 3508 3516
 FT DISULFID 3516 3524
 FT DISULFID 3524 3532
 FT DISULFID 3532 3540
 FT DISULFID 3540 3548
 FT DISULFID 3548 3556
 FT DISULFID 3556 3564
 FT DISULFID 3564 3572
 FT DISULFID 3572 3580
 FT DISULFID 3580 3588
 FT DISULFID 3588 3596
 FT DISULFID 3596 3604
 FT DISULFID 3604 3612
 FT DISULFID 3612 3620
 FT DISULFID 3620 3628
 FT DISULFID 3628 3636
 FT DISULFID 3636 3644
 FT DISULFID 3644 3652
 FT DISULFID 3652 3660
 FT DISULFID 3660 3668
 FT DISULFID 3668 3676
 FT DISULFID 3676 3684
 FT DISULFID 3684 3692
 FT DISULFID 3692 3700
 FT DISULFID 3700 3708
 FT DISULFID 3708 3716
 FT DISULFID 3716 3724
 FT DISULFID 3724 3732
 FT DISULFID 3732 3740
 FT DISULFID 3740 3748
 FT DISULFID 3748 3756
 FT DISULFID 3756 3764
 FT DISULFID 3764 3772
 FT DISULFID 3772 3780
 FT DISULFID 3780 3788
 FT DISULFID 3788 3796
 FT DISULFID 3796 3804
 FT DISULFID 3804 3812
 FT DISULFID 3812 3820
 FT DISULFID 3820 3828
 FT DISULFID 3828 3836
 FT DISULFID 3836 3844
 FT DISULFID 3844 3852
 FT DISULFID 3852 3860
 FT DISULFID 3860 3868
 FT DISULFID 3868 3876
 FT DISULFID 3876 3884
 FT DISULFID 3884 3892
 FT DISULFID 3892 3900
 FT DISULFID 3900 3908
 FT DISULFID 3908 3916
 FT DISULFID 3916 3924
 FT DISULFID 3924 3932
 FT DISULFID 3932 3940
 FT DISULFID 3940 3948
 FT DISULFID 3948 3956
 FT DISULFID 3956 3964
 FT DISULFID 3964 3972
 FT DISULFID 3972 3980
 FT DISULFID 3980 3988
 FT DISULFID 3988 3996
 FT DISULFID 3996 4004
 FT DISULFID 4004 4012
 FT DISULFID 4012 4020
 FT DISULFID 4020 4028
 FT DISULFID 4028 4036
 FT DISULFID 4036 4044
 FT DISULFID 4044 4052
 FT DISULFID 4052 4060
 FT DISULFID 4060 4068
 FT DISULFID 4068 4076
 FT DISULFID 4076 4084
 FT DISULFID 4084 4092
 FT DISULFID 4092 4100
 FT DISULFID 4100 4108
 FT DISULFID 4108 411

Db 2424 VOSSUALSPHAYILGSPSPAKSRPSLPTSTPHIOAMRHATQ-----QKQFGSSNLNSEL 2476
QY 292 -----KNPENOKVN---IDANGN-----AIPSGELXDDIVEQTA 322
Db 2477 GGANGGVVGGGGGGGGVGGQGPQNSPVSLGIISPTGSDWIMLAPPQSSKNSAIMQTIS 2536
QY 323 QOAKEAGEVARQOAVESNAOQRYEDQHARQBEELQSSGIGYGLSSALIYAGGIGAGV 382
Db 2537 PQOQQ-----QOQQOQQOQHOOQQOQQOQQOQQOQQOQQOQQOQQOQQOQQOQQO 2589
QY 383 TTALHRRNQPAEOT-----TTTTTH-----TVVQOQTGG-IPQHKVALM 420
Db 2590 SFHSGQMNPPSTOSSMGSPPSTNMLSPSOHQNOAQFYQVLTPTSSOHSOGGHTPQHLVQIL 2649
QY 421 PQERRRFSRRDSQGSVAETHWSDSSSEYVNPYAE--VGGARNSL-----SAHQPEEHI 472
Db 2650 -----DSYPTSPESPCHWSSSPRSNSDWSGVSQSPAANNLYISGGHQANKGSEAI 2701
QY 473 Y 473
Db 2702 Y 2702

Search completed: September 27, 2001, 14:22:50
Job time: 146 sec

FT DISULFID 85 94 BY SIMILARITY.
FT DISULFID 100 111 BY SIMILARITY.
FT DISULFID 105 124 BY SIMILARITY.
FT DISULFID 126 135 BY SIMILARITY.
FT DISULFID 143 154 BY SIMILARITY.
FT DISULFID 148 164 BY SIMILARITY.
FT DISULFID 166 175 BY SIMILARITY.
FT DISULFID 181 192 BY SIMILARITY.
FT DISULFID 186 203 BY SIMILARITY.
FT DISULFID 205 214 BY SIMILARITY.
FT DISULFID 221 232 BY SIMILARITY.
FT DISULFID 226 241 BY SIMILARITY.
FT DISULFID 243 252 BY SIMILARITY.
FT DISULFID 259 270 BY SIMILARITY.
FT DISULFID 264 279 BY SIMILARITY.
FT DISULFID 281 290 BY SIMILARITY.
FT DISULFID 297 308 BY SIMILARITY.
FT DISULFID 302 317 BY SIMILARITY.
FT DISULFID 319 328 BY SIMILARITY.
FT DISULFID 335 349 BY SIMILARITY.
FT DISULFID 343 358 BY SIMILARITY.
FT DISULFID 360 369 BY SIMILARITY.
FT DISULFID 376 387 BY SIMILARITY.
FT DISULFID 381 396 BY SIMILARITY.
FT DISULFID 398 407 BY SIMILARITY.
FT DISULFID 413 424 BY SIMILARITY.
FT DISULFID 418 435 BY SIMILARITY.
FT DISULFID 437 446 BY SIMILARITY.
FT DISULFID 453 465 BY SIMILARITY.
FT DISULFID 459 474 BY SIMILARITY.
FT DISULFID 476 485 BY SIMILARITY.
FT DISULFID 492 503 BY SIMILARITY.
FT DISULFID 497 512 BY SIMILARITY.
FT DISULFID 514 523 BY SIMILARITY.
FT DISULFID 530 541 BY SIMILARITY.
FT DISULFID 535 550 BY SIMILARITY.
FT DISULFID 552 561 BY SIMILARITY.
FT DISULFID 568 579 BY SIMILARITY.
FT DISULFID 573 588 BY SIMILARITY.
FT DISULFID 590 599 BY SIMILARITY.
FT DISULFID 606 616 BY SIMILARITY.
FT DISULFID 611 625 BY SIMILARITY.
FT DISULFID 627 636 BY SIMILARITY.
FT DISULFID 643 654 BY SIMILARITY.
FT DISULFID 648 663 BY SIMILARITY.
FT DISULFID 665 674 BY SIMILARITY.
FT DISULFID 681 692 BY SIMILARITY.

Query Match 4.6%; Score 130; DB 1; Length 2703;
Best Local Similarity 20.0%; Pred. No. 11;
Matches 120; Conservative 81; Mismatches 194; Indels 206; Gaps 28;

QY 4 GNIGNVNGNHLIPAPPILPSQT--DGAAR-----GCTGHLISSTGALGSRSLFSLRN 55
Db 2177 GGGGNGNGN-----ASGQKQKOTAKQKAKKAKLIEGSPDNGLDATGSLRRK---ASSKK 2229
QY 56 SMADSVDSRDIPGLPTNPRLAAATSETCLLGEFVLHDKGFLDLINTQIGPSAFRVEVQ 115
Db 2230 TSAASKKAANLNL--NFGQLT-----GGVSGVPGVPPTN-----SAAQAAAA 2270
QY 116 ADGTHAAG-ENKGLVSVTLSPQWSSLSQIDTEGKNRFVFTGGRGSGHPWVTVASDI 174
Db 2271 AAAAAMAAASHELEGPVG-----GMGNLPSPYDTSSM 2305
QY 175 AEARTFILAFLPDNHHGROPKDVD-----TRSVGVSASGIDGWSVSETHTSTTNSVR 229
Db 2306 SNAMAAPLANGNP-NTGAKQPPSYEDCTKNAOSMOSLOGNGL-DMIKLDNYAYSMGSPFQ 2363
QY 230 SD-----PKFWVSVGAIAAGLAGL-----AATG----- 252
Db 2364 QELLNCOGLMGNGQNGVPGVLPGLGCGMGLSGAGNGSHSQGLSPPYSNQSPPHS 2423
QY 253 IAQALALTP-----EPDDPTTTDDPQANAAESATKQDLTQBAF----- 291

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 27, 2001, 14:21:49 ; Search time 38.06 Seconds
(without alignments)
1908.446 Million cell updates/sec

Title: US-09-189-415A-2
Perfect score: 2800
Sequence: 1 MPIGNLGNVNGNHLIPPAP.....GETAVSSVNAAPTPGPVRFV 549

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5
Searched: 425026 seqs, 132305027 residues
Total number of hits satisfying chosen parameters: 425026

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_16:
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_unclassified:*
13: sp_vertebrate:*
14: sp_virus:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2795	99.8	549	2	O50190	O50190 escherichia
2	2676.5	95.6	550	2	O52147	O52147 escherichia
3	2496.5	89.2	552	2	O9KWH9	O9KWH9 escherichia
4	2206	78.8	551	2	O68258	O68258 escherichia
5	2077	74.2	547	2	O9WXX1	O9WXX1 escherichia
6	2076	74.1	547	2	O9ETI1	O9ETI1 citrobacter
7	1781.5	63.6	538	2	O47016	O47016 escherichia
8	1781.5	63.6	538	2	O47014	O47014 escherichia
9	1763.5	63.0	538	2	O85508	O85508 escherichia
10	1528.5	54.6	558	2	O9R396	O9R396 escherichia
11	1516.5	54.2	558	2	O85506	O85506 escherichia
12	168.5	6.0	1158	5	O97169	O97169 drosophila
13	160.5	5.7	818	6	O9N1P0	O9N1P0 bos taurus
14	159.5	5.7	1265	5	O9FDA0	O9FDA0 xanthomonas
15	156.5	5.6	1323	5	O9NHX4	O9NHX4 drosophila
16	152	5.4	1203	5	O9N5K0	O9N5K0 caenorhabdi
17	151.5	5.4	2806	2	O9KXA6	O9KXA6 escherichia
18	151.5	5.4	2806	9	O9TK9	O9TK9 bacterioph
19	151	5.4	934	5	O9VN59	O9VN59 drosophila

20	150	5.4	1005	2	O84462	O84462 chl:
21	149.5	5.3	2806	9	O9XJM1	O9XJM1 bac
22	149	5.3	832	2	O54356	O54356 mor
23	149	5.3	2232	5	P91365	P91365 caenorhabdi
24	146	5.2	705	5	O9GYB2	O9GYB2 leishmania
25	146	5.2	1829	5	O22248	O22248 caenorhabdi
26	145.5	5.2	848	5	O9V889	O9V889 drosophila
27	145.5	5.2	2586	5	O9VTK8	O9VTK8 drosophila
28	144.5	5.2	1296	5	O96506	O96506 drosophila
29	144	5.1	1589	6	O62672	O62672 bos taurus
30	143.5	5.1	1778	5	O9NE65	O9NE65 leishmania
31	141.5	5.1	647	3	O14273	O14273 schizosacch
32	140.5	5.0	929	5	O9N993	O9N993 leishmania
33	140.5	5.0	1472	5	O9V358	O9V358 drosophila
34	139	5.0	1064	4	O9HAN2	O9HAN2 homo sapien
35	139	5.0	1860	5	O9UI29	O9UI29 leishmania
36	138.5	4.9	2353	2	P71401	P71401 haemophilus
37	138	4.9	1455	5	O9VP05	O9VP05 drosophila
38	138	4.9	2857	5	O9VXZ1	O9VXZ1 drosophila
39	137	4.9	1066	11	O9ERC7	O9ERC7 mus musculu
40	137	4.9	1475	5	O9GYW7	O9GYW7 drosophila
41	137	4.9	2468	2	O9I2M3	O9I2M3 pseudomonas
42	136.5	4.9	770	3	O9P3I9	O9P3I9 neurospora
43	136.5	4.9	1829	2	O86808	O86808 streptomyce
44	136	4.9	540	9	O9G0H8	O9G0H8 roseophage
45	136	4.9	674	2	O9KQ73	O9KQ73 vibrio chol

ALIGNMENTS

RESULT	1
O50190	
ID	O50190 PRELIMINARY; PRT; 549 AA.
AC	O50190;
DT	01-JUN-1998 (TREMREL. 06, Created)
DT	01-JUN-1998 (TREMREL. 06, Last sequence update)
DT	01-MAR-2001 (TREMREL. 16, Last annotation update)
DE	TRANSLOCATED INTIMIN RECEPTOR.
GN	TIR.
OS	Escherichia coli.
OC	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC	Escherichia.
OX	NCBI_TaxID=562;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=E2348/69;
RX	MEDLINE=98050926; PubMed=9390560;
RA	Kenny B., Devlinney R., Stein M., Reinscheid D.J., Frey E.A.,
RA	Finlay B.B.;
RT	"Enteropathogenic E. coli (EPEC) transfers its receptor for intimate
RT	adherence into mammalian cells.";
RL	Cell 91:511-520(1997).
DR	EMBL; AF013122; AAB88410.1; -
DR	InterPro; IPR003536; -
DR	PRINTS; PR01370; TRNSINTMINR.
SQ	SEQUENCE 549 AA; 56843 MW; 40C8B8B234409A08 CRC64;

Query Match 99.8%; Score 2795; DB 2; Length 549;
Best Local Similarity 99.6%; Pred. No. 6.4e-155;
Matches 547; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY	1	MPIGNLGNVNGNHLIPPAPLPSTQDGAAGTGTHLISSTGALGSRSLFSPLRNSMADS	60
Db	1	MPIGNLGNVNGNHLIPPAPLPSTQDGAAGTGTHLISSTGALGSRSLFSPLRNSMADS	60
QY	61	VPSRDIPLGLTPNPSRLAAATSETCLLGGEVLHDKGPLDLNTQIGPSAFRVEQADGTH	120
Db	61	VPSRDIPLGLTPNPSRLAAATSETCLLGGEVLHDKGPLDLNTQIGPSAFRVEQADGTH	120
QY	121	AAIGEKNGLEVSTLSPQEWSSLSQSDTECKNRFVFTGGGSGGHPMTVAVSDIAEAPTR	180
Db	121	AAIGEKNGLEVSTLSPQEWSSLSQSDTECKNRFVFTGGGSGGHPMTVAVSDIAEAPTR	180

```
Db 121 AAIGKNGLEVSVTLSPQEWSSLSQSDTEGKNRFVFTGGRGSGHPMTVTVASDIAEARTK 180
Qy 181 ILAKLDPDNHGGQPKDVTDRSVGVSGASGIDGCVVSETHSTTNSSVRSDPKFWVSVGA 240
Db 181 ILAKLDPDNHGGQPKDVTDRSVGVSGASGIDGCVVSETHSTTNSSVRSDPKFWVSVGA 240
Qy 241 IAAGLAGLAATGIAQALALTPEDDPTTDPDQAANAESATKDQLTQEAFAKPNQKVN 300
Db 241 IAAGLAGLAATGIAQALALTPEDDPTTDPDQAANAESATKDQLTQEAFAKPNQKVN 300
Qy 301 IDANGNAIPSGELXDDIVEQIAQAQKEAGEVARQQAQAVESNAQAQRYEDQHARRQELQL 360
Db 301 IDANGNAIPSGELXDDIVEQIAQAQKEAGEVARQQAQAVESNAQAQRYEDQHARRQELQL 360
Qy 361 IDANGNAIPSGELXDDIVEQIAQAQKEAGEVARQQAQAVESNAQAQRYEDQHARRQELQL 360
Db 361 IDANGNAIPSGELXDDIVEQIAQAQKEAGEVARQQAQAVESNAQAQRYEDQHARRQELQL 360
Qy 421 POERRRFSDDRRDSQGSVASTHWSDSSEVVNPNYAEVGGARNSLSAHOPEEHYDEVAADP 480
Db 421 POERRRFSDDRRDSQGSVASTHWSDSSEVVNPNYAEVGGARNSLSAHOPEEHYDEVAADP 480
Qy 481 GYSVTQNFSGSPVTRGLIGTPGQGIQSTYALLANSGLRLGMLTSGGETAVSSVNA 540
Db 481 GYSVTQNFSGSPVTRGLIGTPGQGIQSTYALLANSGLRLGMLTSGGETAVSSVNA 540
Qy 541 PTPGPVRFV 549
Db 541 PTPGPVRFV 549

RESULT 2
O52147 PRELIMINARY; PRT; 550 AA.
AC O52147;
DT 01-JUN-1998 (Tremblrel. 06, Created)
DT 01-JUN-1998 (Tremblrel. 06, Last sequence update)
DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)
DE TRANSLOCATED INTIMIN RECEPTOR.
GN TIR.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RA Elliott S.J., Wainwright L.A., McDaniel T.K., Jarvis K.G., Deng Y.K.,
RA Lai L.C., McNamara B.P., Donnenberg M.S., Kaper J.B.;
RL Mol. Microbiol. 0:0-0(1998).
DR EMBL; AF022236; AAC38390.1; -.
DR InterPro; IPR003536; -.
DR PRINTS; PR01370; TRNSINTMINR.
SQ SEQUENCE 550 AA; 56509 MW; 19DD08A9BE9251CB CRC64;

Query Match 95.6%; Score 2676.5; DB 2; Length 550;
Best Local Similarity 96.0%; Pred. No. 5e-148;
Matches 531; Conservative 4; Mismatches 11; Indels 7; Gaps 2;

Qy 1 MPIGNLGNVNGNHLIPAPPPLPSQTDGAARGGTGHLISSTGALGSRSLFSPLRNSMADS 60
Db 1 MPIGNLGNVNGNHLIPAPPPLPSQTDGAARGGTGHLISSTGALGSRSLFSPLRNSMADS 60
Qy 61 VDSRDIPGLTPNPSRLAAATSETCLGGFEVLHDKGPDILNTQIGPSAFRVEVQADGTH 120
Db 61 VDSRDIPGLTPNPSRLAAATSETCLGGFEVLHDKGPDILNTQIGPSAFRVEVQADGTH 120
Qy 121 AATGEKNGLEVSVTLSPQEWSSLSQSDTEGKNRFVFTGGRGSGHPMTVTVASDIAEARTK 180
Db 121 AATGEKNGLEVSVTLSPQEWSSLSQSDTEGKNRFVFTGGRGSGHPMTVTVASDIAEARTK 180
Qy 181 ILAKLDPDNHGGQPKDVTDRSVGVSGASGIDGCVVSETHSTTNSSVRSDPKFWVSVGA 240
Db 181 ILAKLDPDNHGGQPKDVTDRSVGVSGASGIDGCVVSETHSTTNSSVRSDPKFWVSVGA 240
```

```
Db 181 ILAKLDPDNHGGQPKDVTDRSVGVSGASGIDGCVVSETHSTTNSSVRSDPKFWVSVGA 240
Qy 241 IAAGLAGLAATGIAQALALTPEDDPTTDPDQAANAESATKDQLTQEAFAKPNQKVN 300
Db 241 IAAGLAGLAATGIAQALALTPEDDPTTDPDQAANAESATKDQLTQEAFAKPNQKVN 300
Qy 301 IDANGNAIPSGELXDDIVEQIAQAQKEAGEVARQQAQAVESNAQAQRYEDQHARRQELQL 360
Db 301 IDANGNAIPSGELXDDIVEQIAQAQKEAGEVARQQAQAVESNAQAQRYEDQHARRQELQL 360
Qy 361 SSIGYGLSSALIVAGGIGAGVTALHRRNOPAEQTTTTTHTVVQOQTGGIPQHKVAM 420
Db 361 SSIGYGLSSALIVAGGIGAGVTALHRRNOPAEQTTTTTHTVVQOQTGGIPQHKVAM 420
Qy 421 POERRRFSDDRRDSQGSVASTHWSDSSEVVNPNYAEVGGARNSLSAHOPEEHYDEV 476
Db 421 GTDATRAEDASLNRDSQGSVASTHWSDSSEVVNPNYAEVGGARNSLSAHOPEEHYDEV 477
Qy 477 AADPGYVIONFSGSPVTRGLIGTPGQGIQSTYALLANSGLRLGMLTSGGETAVSS 536
Db 478 AADPGYVIONFSGSPVTRGLIGTPGQGIQSTYALLANSGLRLGMLTSGGETAVSS 537
Qy 537 VNAAPTGPVRFV 549
Db 538 VNAAPTGPVRFV 550

RESULT 3
Q9KWH9 PRELIMINARY; PRT; 552 AA.
AC Q9KWH9;
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)
DE TIR.
GN TIR.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RA Abe A., Nagano H.;
RA "Analyses of type III secreted proteins and Tir in enteropathogenic
RT Escherichia coli O157:H45.";
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB036053; BAA96815.1; -.
DR InterPro; IPR003536; -.
DR PRINTS; PR01370; TRNSINTMINR.
SQ SEQUENCE 552 AA; 57005 MW; A8D79EE22EE50A4B CRC64;

Query Match 89.2%; Score 2496.5; DB 2; Length 552;
Best Local Similarity 90.3%; Pred. No. 1.4e-137;
Matches 501; Conservative 11; Mismatches 34; Indels 9; Gaps 3;

Qy 1 MPIGNLGNVNGNHLIPAPPPLPSQTDGAARGGTGHLISSTGALGSRSLFSPLRNSMADS 60
Db 1 MPIGNLGNVNGNHLIPAPPPLPSQTDGAARGGTGHLISSTGALGSRSLFSPLRNSMADS 60
Qy 61 VDSRDIPGLTPNPSRLAAATSETCLGGFEVLHDKGPDILNTQIGPSAFRVEVQADGTH 120
Db 61 VDSRDIPGLTPNPSRLAAATSETCLGGFEVLHDKGPDILNTQIGPSAFRVEVQADGTH 120
Qy 121 AATGEKNGLEVSVTLSPQEWSSLSQSDTEGKNRFVFTGGRGSGHPMTVTVASDIAEARTK 180
Db 121 AATGEKNGLEVSVTLSPQEWSSLSQSDTEGKNRFVFTGGRGSGHPMTVTVASDIAEARTK 180
Qy 181 ILAKLDPDNHGGQPKDVTDRSVGVSGASGIDGCVVSETHSTTNSSVRSDPKFWVSVGA 240
Db 181 ILAKLDPDNHGGQPKDVTDRSVGVSGASGIDGCVVSETHSTTNSSVRSDPKFWVSVGA 240
```

QY 241 IAAGLAGLAATGIAQALALTPEPDDPTTTTDPQAAANAESATKDQLTQAFKNPNQKYN 300
|||||
Db 241 IAAGLAGLAATGIAQALALTPEPDDPTTTTDPQAAANAESATKDQLTQAFKNPNQKYN 300
QY 301 IDANGNAIPSGELXDDIVBQIAQAQAKEAGEVARQAQAVESNAQAQRYEDQHARROEELQ 360
|||||
Db 301 IDANGNAIPSGELXDDIVBQIAQAQAKEAGEVARQAQAVESNAQAQRYEDQHARROEELQ 360
QY 361 SSGIGYGLSSALIVAGGIGAGVTTTALHRRNQPAEQTTTTTTHVVVQOQTGGIPQHKVALM 420
|||||
Db 361 SSGIGYGLSSALIVAGGIGAGVTTTALHRRNQPAEQTTTTTTHVVVQOQTGGIPQHKVALM 420
QY 361 SSGIGYGLSSALIVAGGIGAGVTTTALHRRNQPAEQTTTTTTHVVVQOQTGGIPQHKVALM 420
Db 361 SSGIGYGLSSALIVAGGIGAGVTTTALHRRNQPAEQTTTTTTHVVVQOQTGGIPQHKVALM 420
QY 421 POERRRFS- ---RDSQGSVASTHWSDSSEVVNPYAEVGGARNSLSAHQPEEHYDEV 476
|||||
Db 421 POERRRFS- ---RDSQGSVASTHWSDSSEVVNPYAEVGGARNSLSAHQPEEHYDEV 476
QY 477 AADPGYSVLQNSGSGPVTVGRLLIGTPGQGIQSTYALLA-NSGGLRLGMGLTSGGETAV-- 534
|||||
Db 477 AADPGYSVLQNSGSGPVTVGRLLIGTPGQGIQSTYALLA-NSGGLRLGMGLTSGGETAV-- 534
QY 535 SSVNAAPTGPVRFV 549
::: |||||
Db 538 ANANAAPTGPVRFV 552

RESULT 4
O68258 PRELIMINARY; PRT; 551 AA.
AC O68258
DT 01-AUG-1998 (TREMBLrel. 07, Created)
DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE TRANSLOCATED INTIMIN RECEPTOR.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98187918; PubMed=9529069;
RA Voss E., Paton A.W., Manning P.A., Paton J.C.;
RT "Molecular analysis of Shiga toxinigenic Escherichia coli O111:H-
RT proteins which react with sera from patients with hemolytic-uremic
RT syndrome";
RL Infect. Immun. 66:1467-1472(1998).
DR EMBL; AF025311; AAC69249.1; -;
DR InterPro; IPR003536; -;
DR PRINTS; PRO1370; TRNSINTIMINR.
SQ SEQUENCE 551 AA; 56975 MW; 6EC95F76E0F44CC CRC64;

Query Match 78.8%; Score 2206; DB 2; Length 551;
Best Local Similarity 79.8%; Pred. No. 1.1e-120;
Matches 442; Conservative 39; Mismatches 65; Indels 8; Gaps 3;
QY 1 MPIGNLGNVNNHILPPAPPPLPSQTDGAARGTGHLISSTGALGSRSLFSPLRNSMADS 60
|||||
Db 1 MPIGNLGNVNNHILPPAPPPLPSQTDGAARGTGHLISSTGALGSRSLFSPLRNSMADS 60
QY 61 VDSRDIPGLPTNPRLAAATSETCLLGGFEVLHDKGPLDLINTQIGPSAFRVEVQADGTH 120
|||||
Db 61 VDSRDIPGLPVHPRLATATSEICLLGGFEVLHDKGPLDLNTKQIGASAFRIEQQSDGSY 120
QY 121 AAIKENGLEVSVTLSPOQWSSLOSIDTEGKNRFVFTGGGSGHPMTVVASDIAEARTR 180
|||||
Db 121 AAIKENGLEVSVTLSPOQWSSLOSIDTEGKNRFVFTGGGSGHPMTVVASDIAEARAK 180
QY 181 ILAKLDPDNHGGROPKDVTRSVGVSAGSIDGVVSETHSTTNSSVRSRDPKFWYSVGA 240
|||||
Db 181 ILAKLDPDNHGGSQARNVTRSVGVSAGSMDDSVVSETHSTTNSSVRSRDPKFWYSVGA 240
QY 241 IAAGLAGLAATGIAQALALTPEPDDPTTTTDPQAAANAESATKDQLTQAFKNPNQKYN 300
|||||

Db 241 IAAGLAGLAATGIAQALALTPEPDDPTTTTDPQAAANAESATKDQLTQAFKNPNQKYS 300
QY 301 IDANGNAIPSGELXDDIVBQIAQAQAKEAGEVARQAQAVESNAQAQRYEDQHARROEELQ 360
|||||
Db 301 IDEIGNSIPSGELKDDVAKIEQAKEAGEAARQAQAVESNAQAQRYDYQYARRQEEL 360
QY 361 SSGIGYGLSSALIVAGGIGAGVTTTALHRRNQPAEQTTTTTTHVVVQOQTGGIPQHKVALM 420
|||||
Db 361 SSGIGYGLSSALIVAGGIGAGVTTTALHRRNQPAEQTTTTTTHVVVQOQTGGIPQHKVALM 420
QY 421 POERRRFS- ---RDSQGSVASTHWSDSSEVVNPYAEVGGARNSLSAHQPEEHYDEV 476
|||||
Db 418 GTDATRAEDASLNRDQSQRSTASTHWSDTSSAVNPYAEVGEARNSSPARQAEHYDEV 477
QY 477 AADPGYSVLQNSGSGPVTVGRLLIGTPGQGIQSTYALLA-NSGGLRLGMGLTSGGETAVS 535
|||||
Db 478 AADPNYSVLQNSGSGPVTVGRLLIGTPGQGIQSTYALLTNNNSAGLRLGMGLTSGGSAYN 537
QY 536 SYNAAPTGPVRFV 549
: |||||
Db 538 TANAAPTGPGRFV 551
RESULT 5
Q9WXK1 PRELIMINARY; PRT; 547 AA.
AC Q9WXK1
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE TRANSLOCATED INTIMIN RECEPTOR.
GN TIR.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MPEC;
RA Okutani A., Itoh K., Sasakawa C.;
RT "Translocated intimin receptor (Tir) of murine pathogenic Escherichia
RT coli O15a,c:K(B).";
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB026719; BAA77400.1; -;
DR InterPro; IPR000484; -;
DR InterPro; IPR003536; -;
DR PRINTS; PRO1370; TRNSINTIMINR.
DR ProDom; PD000551; -; 1.
KW Receptor.
SQ SEQUENCE 547 AA; 56283 MW; 02CAC6D625FA6EE1 CRC64;

Query Match 74.2%; Score 2077; DB 2; Length 547;
Best Local Similarity 76.0%; Pred. No. 3.5e-113;
Matches 418; Conservative 44; Mismatches 84; Indels 4; Gaps 3;
QY 1 MPIGNLGNVNNHILPPAPPPLPSQTDGAARGTGHLISSTGALGSRSLFSPLRNSMADS 60
|||||
Db 1 MPIGNLGNVNNHILPPAPPPLPSQTDGAARGTGHLISSTGALGSRSLFSPLRNSMADS 60
QY 61 VDSRDIPGLPTNPRLAAATSETCLLGGFEVLHDKGPLDLINTQIGPSAFRVEVQADGTH 120
|||||
Db 61 VDSRDVPGLPEHPLRF-ATSETCLHGGFEVLHDKGPLDLNTKQIGSVFRVEVQPDGTH 118
QY 121 AAIKENGLEVSVTLSPOQWSSLOSIDTEGKNRFVFTGGGSGHPMTVVASDIAEARTR 180
|||||
Db 119 AAIKVGKDGVEVSVTLNSSLQSLQSLDTGNGRFVFTGGGSGHPMTVVASDIAEQAREK 178
QY 181 ILAKLDPDNHGGROPKDVTRSVGVSAGSIDGVVSETHSTTNSSVRSRDPKFWYSVGA 240
|||||
Db 179 ILAKLDPDNHGGROPKDVTRSVGVSAGSMDDSVVSETHSTTNSSVRSRDPKFWYSVGA 238
QY 241 IAAGLAGLAATGIAQALALTPEPDDPTTTTDPQAAANAESATKDQLTQAFKNPNQKYN 300
|||||

```

|||||
239 IAAGLAGLAATGIVQAVALTAPDDPTTDPDEAANAEEAATKDQLTKEAFONPDNQVN 298
QY 301 IDANGNAIPSGELXDDIVEQIAQAQKEAGEVARQAQVESAQAQRYEDQHARRQEELQL 360
Db 299 IDELGNAPSGELKDDVVAQIADQAKVAGEARQAQVESAQAQRRHDDQQAQROEQLDL 358
QY 361 SSGIGYGLSSALIVAGGIGAGVTTALHRRNOPAEQTTTTTTHVVQOQTGGIPQHKVALM 420
Db 359 SSGIGYGLSSALIVAGGIGAGVTTALHRRNPTEQ-TIATTHSVIOOQTGGNTRAOGGAD 417
QY 421 PQERRRES-DRRDSQGSVASTHWSDSSEVWNPYAEVGGARNSLSAHQPEEHYDEVAAD 479
Db 418 TTGVENASLTRRDSQASVASTQSDTSGDVVNPYABGWSRNPSPSLAPEEPIYDEVAPD 477
QY 480 PGYSVIONFSGSPVTRGLIGTPGQIQSTYALLANSGLRLGMGLTSGGETAVSSVNA 539
Db 478 PNYSVIOHFSGNPNVTGRVLGSPGQIQSTYALLANSGLRLGMGLTSGGESAGSAANA 537
QY 540 APTPGPVREFV 549
Db 538 ATTPGVREFV 547

```

```

RESULT 7
Q47016 PRELIMINARY; PRT; 538 AA.
ID Q47016;
AC Q47016;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-AUG-1999 (TREMBlrel. 11, Last sequence update)
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
DE TRANSLOCATED INTIMIN RECEPTOR.
GN TIR.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-REPEC 83/39;
RA O'Brien R.A.;
RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.
[2]
RP SEQUENCE FROM N.A.
RC STRAIN-REPEC 83/39;
RA Krejany E.O.;
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
[3]
RP SEQUENCE FROM N.A.
RC STRAIN-RDEC-1;
RA Agin T.S., Boedeker E.C.;
RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
[4]
RP SEQUENCE OF 172-538 FROM N.A.
RC STRAIN-REPEC RDEC-1;
RA Krejany E.O.;
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; U59504; AAD19750.1; -
DR EMBL; AF045568; AAC15683.1; -
DR EMBL; U59503; AAB02941.1; -
DR InterPro; IPR003536; -
DR PRINTS; PR01370; TRNSINTMINR.
KW Receptor.
SQ SEQUENCE 538 AA; 55410 MW; 3132A969B7B3D06C CRC64;

```

```

Query Match 63.6%; Score 1781.5; DB 2; Length 538;
Best Local Similarity 65.4%; Pred. No. 5.1e-96;
Matches 367; Conservative 66; Mismatches 93; Indels 35; Gaps 11;
QY 1 MPIGNLGNVGNHILIPAPPPLPSQTDGAARCGTGHLLISSTGALGSRSLFSPLRNSMADS 60
Db 1 MPIGNLGNHNSRALIPAPPPLPSQTDGAG-GARNQLINSNGPMGSRLLFTFIRNSVADA 59

```

```

|||||
239 IAAGLAGLAATGIVQAVALTAPDDPTTDPDEAANAEEAATKDQLTKEAFONPDNQVN 298
QY 301 IDANGNAIPSGELXDDIVEQIAQAQKEAGEVARQAQVESAQAQRYEDQHARRQEELQL 360
Db 299 IDELGNAPSGELKDDVVAQIADQAKVAGEARQAQVESAQAQRRHDDQQAQROEQLDL 358
QY 361 SSGIGYGLSSALIVAGGIGAGVTTALHRRNOPAEQTTTTTTHVVQOQTGGIPQHKVALM 420
Db 359 SSGIGYGLSSALIVAGGIGAGVTTALHRRNPTEQ-TIATTHSVIOOQTGGNTRAOGGAD 417
QY 421 PQERRRES-DRRDSQGSVASTHWSDSSEVWNPYAEVGGARNSLSAHQPEEHYDEVAAD 479
Db 418 TTGVENASLTRRDSQASVASTQSDTSGDVVNPYABGWSRNPSPSLAPEEPIYDEVAPD 477
QY 480 PGYSVIONFSGSPVTRGLIGTPGQIQSTYALLANSGLRLGMGLTSGGETAVSSVNA 539
Db 478 PNYSVIOHFSGNPNVTGRVLGSPGQIQSTYALLANSGLRLGMGLTSGGESAGSAANA 537
QY 540 APTPGPVREFV 549
Db 538 ATTPGVREFV 547

```

```

RESULT 6
Q9ET11 PRELIMINARY; PRT; 547 AA.
ID Q9ET11;
AC Q9ET11;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
DE TRANSLOCATED INTIMIN RECEPTOR TIR.
OS Citrobacter rodentium.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Citrobacter.
OX NCBI_TaxID=67825;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CDC 1843-73T, AND DBS100;
RX MEDLINE-2055330; PubMed-11101562;
RA Luperchio S.A., Newman J.V., Dangler C.A., Schrenzel M.D.,
RA Brenner D.J., Steigerwalt A.G., Schauer D.B.;
RT "Citrobacter rodentium, the Causative Agent of Transmissible Murine
RT Colonic Hyperplasia, Exhibits Clonality: Synonymy of C. rodentium and
RT Mouse-Pathogenic Escherichia coli.";
RL J. Clin. Microbiol. 38:4343-4350(2000).
DR EMBL; AF301618; AAG40758.1; -
DR EMBL; AF301617; AAG25642.1; -
KW Receptor.
SQ SEQUENCE 547 AA; 56270 MW; C88318B301049C37 CRC64;

```

```

Query Match 74.1%; Score 2076; DB 2; Length 547;
Best Local Similarity 76.0%; Pred. No. 4e-113;
Matches 418; Conservative 44; Mismatches 84; Indels 4; Gaps 3;
QY 1 MPIGNLGNVGNHILIPAPPPLPSQTDGAARCGTGHLLISSTGALGSRSLFSPLRNSMADS 60
Db 1 MPIGNLGNHNSRALIPAPPPLPSQTDGATRGSGSLISSTGSLRSLFSPLRNSIVDT 60
QY 61 VDSRDIPGLTPNPSRLAATSETCLLGGFEVLHDKGPDILNTQIGPSAFRVEQADGTH 120
Db 61 VDSRDVPGLEPHLRF--ATSETCLHGGFEVLHDKGPDILNTQIGSSVFRVEQPDGTH 118
QY 121 AATGKNGLEVSTLSPOEWSLSQSDTEGKRFVFTGGGSGHPMTVVASDIAEARTR 180
Db 119 AATGVKDGVEVSTLSNSELQSLSDTETGTRFTGTGGGSGHAMVTVVASDIAEQAREK 178
QY 181 ILAKLPDNGHGRQKDVTRSVGVSAGSIDDVGVVSETHSTTNNSSVSRSDPKFWVSVGA 240
Db 179 ILAKLPDNGHGRQKDVTRSVGVSAGSGMDGVVSETHSTTNNSSVSRSDPKFWVSVGA 238
QY 241 IAAGLAGLAATGIAQALATPEPDDPTTDPQANAEEAATKDQLTKEAFONPDNQVN 300

```


DE	TRANSLOCATED INTIMIN RECEPTOR TIR.	DE	TRANSLOCATED INTIMIN RECEPTOR TIR.
GN	TIR.	GN	TIR.
OS	Escherichia coli.	OS	Escherichia coli.
OC	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;	OC	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC	Escherichia.	OC	Escherichia.
OX	NCBI_TaxID=562;	OX	NCBI_TaxID=562;
RN	[1]	RN	[1]
RP	SEQUENCE FROM N.A.	RP	SEQUENCE FROM N.A.
RC	STRAIN=EP687A;	RC	STRAIN=EP687A;
RA	Paton A.W., Manning P.A., Woodrow M.C., Paton J.C.:	RA	Perna N.T., Mayhew G.F., Posfal G., Elliott S., Donnenberg M.S.,
RT	"The translocated intimin receptors (Tir) of Shiga toxinigenic	RT	Kaper J.B., Blattner F.R.;
RT	Escherichia coli isolates belonging to serogroups O26, O111, and O157	RT	"Molecular evolution of a pathogenicity island from enterohemorrhagic
RT	react with sera from patients with hemolytic-uremic syndrome and	RL	Escherichia coli O157:H7.;"
RT	exhibit marked sequence heterogeneity.;"	RL	Infect. Immun. 66:3810-3817(1998).
RL	Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.	DR	EMBL; AF125993; AAC29391.1; -
DR	EMBL; AF070069; AAC69318.1; -	DR	InterPro; IPR003536; -
DR	InterPro; IPR003536; -	DR	PRINTS; PR01370; TRNSINTIMINR.
DR	PRINTS; PR01370; TRNSINTIMINR.	KW	Receptor.
SQ	SEQUENCE 538 AA; 55602 MW; 447052A0E3214D6D CRC64;	SQ	SEQUENCE 558 AA; 58022 MW; 99C417222D4B4AA1 CRC64;
Query Match 63.0%; Score 1763.5; DB 2; Length 538;		Query Match 54.6%; Score 1528.5; DB 2; Length 558;	
Best Local Similarity 64.5%; Pred. No. 5.7e-95; Indels 35; Gaps 11;		Best Local Similarity 56.2%; Pred. No. 2.7e-81;	
Matches 362; Conservative 71; Mismatches 93;		Matches 327; Conservative 60; Mismatches 138; Indels 57; Gaps 11;	
QY	1 MPICNLNNVNGNHLIPAPPPLPSQTDGAARGGTGHLISSTGALSGSLSPRLNSMADS 60	QY	1 MPICNLNNVNGNHLIPAPPPLPSQTDGAARGGTGHLISSTGALSGSLSPRLNSMADS 60
DB	1 MPICNLGNHNPVRALIPAPPPLPSQTDGAG-GARNQLNSNGPMGSRLLTPTIRNSVADA 59	DB	1 MPICNLGNHNPVNNISIPAPPPLPSQTDGA--GGRGLINSTGPLGSRALFTPVNSMADS 58
QY	61 VDSR--DIPGLTPNPSRLAAATSETCLLGGFEVLHDKGPDILNTQIGPSAFRVEQADG 118	QY	61 VDSR--DIPGLTPNPSRLAAATSETCLLGGFEVLHDKGPDILNTQIGPSAFRVEQADG 118
DB	60 ADSRASDIPGLTPNPLFAA--SEVSLHGALEVLHDKGGLDTLNSAIGSSLFVRVETRDDG 117	DB	59 GDNRASDPGLPVNPMRLAA--SEITLNDGFEVLHDKGPDILNTQIGSSVFRVETQEDG 116
QY	119 THAATGKNGLEVSVTLSPQEWSSLSQSDTEGKNRFVFTGGRGSGHPMTVASDIAEAR 178	QY	119 THAATGKNGLEVSVTLSPQEWSSLSQSDTEGKNRFVFTGGRGSGHPMTVASDIAEAR 178
DB	118 SHVATGKNGLEVSVTLSPQEWSSLSQSDTEGKNRFVFTGGRGSGHPMTVASDIAEAR 177	DB	117 KHIAVGQRNGVETSVVLSQDYARLQSLDPGKDFVFTGGRGSGHPMTVASDIAEAR 176
QY	179 TRIAKLDP--DNHGGQRPKVDYTRSVGVSASGIDGVSSTHTSTNSVRSRDPKFWVS 237	QY	179 TRIAKLDP--DNHGGQRPKVDYTRSVGVSASGIDGVSSTHTSTNSVRSR 230
DB	178 QRIIDKLEPKDTKETKEGDPNS-----GEGKIIIEHTSTSTFLSRADPKLWLS 226	DB	177 QRILLEPEKGTG-----ESKGAGESKGVGELRESNSGAENTTETOTSTSSLSRS 227
QY	238 VCAIAGLAGLAATGIAQAALATPEPDPTTDPDOAANAASATKDOLTQEAFAKNPENQ 297	QY	231 DPKFWVSGATAAGLAGLAATGIAQAALATPEPDPTTDPDOAANAASATKDOLTQEA 290
DB	227 LGTIAAGLIGMAATGIAQAALATPEPDPTTDPDOAANAASATKDOLTQEAFAKNPENQ 286	DB	228 DPKLWALGTATGLIGLAATGIAQAALATPEPDPTTDPDOAANAASATKDOLTQEA 287
QY	298 KVNIDANGNAIPSGELXDDIVEQIAQAQAGEVARQAQAVESNAQAQORYEDQHARQEE 357	QY	291 FKNPNQKVNIDANGNAIPSGELXDDIVEQIAQAQAGEVARQAQAVESNAQAQORYEDQ 350
DB	287 KVNIDANGNAIPSGELIDVVVAQIAQAQAGEVARQAQAVESNAQAQORYEDQHARQEE 346	DB	288 FQNPQKVNIDANGNAIPSGELXDDIVEQIAQAQAGEVARQAQAVESNAQAQORYEDQ 347
QY	358 LQLSSGIGYGLSSALIVAGGIGAGVTTALHRRNQPAEQTTTTHTTVVQQGTGGIPQHKV 417	QY	351 HARROEELQLSSGIGYGLSSALIVAGGIGAGVTTALHRRNQPAEQTTTTHTTVV 403
DB	347 MSLSGIGYGLSSALIVAGGIGAGVTTALHRRNQPAEQTTTTHTTVVVDNQ-----PTNNA 400	DB	348 QAKROEELKVSAGYGLSGALILGGIGAGVTTALHRRNQPAEQTTTTHTTVV 407
QY	418 ALM-----PQERRFRDRDSQGSVASTHWSDS--EVNPNYAEGVGNLSAHQP 468	QY	404 VYQOQTGGIPQHKVLMQERRFRDRDSQGSVASTHWSDS--EVNPNYAEGVGNLS 462
DB	401 SAQNGNTDTSGPESP--ASRRNSNLSAS--NGSDTSSITGVENPADVGMPRNDSLARIP 457	DB	408 VENKPNANTPAGNVVDTPGSEDTMESRRSSMASTSTFTDTSIGTVQNPYADV-----K 462
QY	469 EEHYDEVAADPGYVIQNFSGSGPVGTGLIGTPQGGIQTSTYALLANGSGLRLGMGLTGS 528	QY	463 LSAHQPEEHIYDEVAADPGYVIQNFSGSGPVGT-----GRLTGTGCGTQ 507
DB	458 EEPIYDEVAADPNYSVIOHFSGNSPTVGLVGTGPGGIGTSTYALLANGSGLRLGMGLTGS 517	DB	463 TSLHD-----SQVPTSNSNTSVQNMGNTPDSVYVSTIQHPDRDTPDNGARLLGNPSAGIQ 516
QY	529 GGTAVSSVNAAPTGPVRFV 549		
DB	518 GGSASVSTANASPTGPVRFV 538		
RESULT	10		
Q9R396	PRELIMINARY; PRT: 558 AA.		
ID	Q9R396		
AC	Q9R396		
DT	01-MAY-2000 (TrEMBLrel. 13, Created)		
DT	01-MAY-2000 (TrEMBLrel. 13, Last sequence update)		
DT	01-MAR-2001 (TrEMBLrel. 16, Last annotation update)		


```
Db 540 TADRISSNGGGGAGISL--SHSLAHKVPSPSSAAAASRLVYHHHHQHVSPKRILREF 597
QY 474 DEVAADPGYGVIONFNGSGPVTGRLICTPGQGIQSTYALLANSGLRGMGLTS--GGE 531
Db 598 EKVSLIEDNNGCVNNGSGV-----GGISGGAGGKRRAKGS 633
QY 532 TAVSSV-----NAAP-----TPGP 545
Db 634 TATSAVTTKSMPINLAPPAQAKVLSTPTP 661

RESULT 13
Q9N1P0 PRELIMINARY; PRT; 818 AA.
AC Q9N1P0;
DT 01-OCT-2000 (T-EMBLrel. 15, Created)
DT 01-OCT-2000 (T-EMBLrel. 15, Last sequence update)
DT 01-MAR-2001 (T-EMBLrel. 16, Last annotation update)
DE SUBMAXILIARY MUCIN (FRAGMENT).
GN BSM1.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BREED ANGUS;
RA MEDLINE=2023253; PubMed=10759843;
RX Jang W., Gupta D., Gallagher D., Davis S., Bhavanandan V.P.;
RT "The central domain of bovine submaxillary mucin consists of over 50
RT tandem repeats of 329 amino acids: chromosomal localization of the
RT BSM1 gene and relations to ovine and porcine counterparts.";
RL Eur. J. Biochem. 267:2208-2217(2000).
DR EMBL: AF178428; AAF67279.1; -.
FT NON_TER 818
FT NON_TER 818
SQ SEQUENCE 818 AA; 73014 MW; B5BB44F84F66F86B CRC64;
```

```
Query Match 5.7%; Score 160.5; DB 6; Length 818;
Best Local Similarity 21.68; Pred. No. 0.15; Mismatches 221; Indels 157; Gaps 21;
Matches 122; Conservative 64;

QY 10 VGNHLPAPPPLPSQTDGAARGTGHLISSTGALGSRSLFSPLRNMSDVSDDIPGL 69
Db 174 VSGRAVPGSPGRSRATAVSGEGSQPTVALSGATGTSAGPSGRST-----SSAIPAT 227
QY 70 PTNPSRLAAATSETCLLGGFEVLHDKGPDILNTQIGPSAFRVEQDCTHAATGEKNGL 129
Db 228 PGSTTGAAGAGTGA-----VDSQQTASLPAAARPTALGPSTAPSGTSES 274
QY 130 EVSVTLSPQWSSLSQSDTEGKRFVFTGGRGGHGMVTVASDIAEARTI---LAKL 185
Db 275 RSSVP-----GGSETTQPG-----AGSEPTLSPGVTTTALRGSETRVPSTGVSGL 322
QY 186 DPNHGGRPQKVDTRSVGVSAGSIDGVVSETHST---TNSSVSRDPPKFWVSVGAIA 242
Db 323 PGSTGG-----SAATGGAGSGPTAPVSGEWTSTVISGTNPVSGAP---VTPGSSA 373
QY 243 -----AGLAGLATGIAQALALTPEDDPTTDDQANAAESATK 283
Db 374 GSSGAPGTGGPGSETASPLSGAAGTSATGSGTSI---PPSGAPVTPPEPLISTGA-SAGP 429
QY 284 DQLTQEAFFKPNQKVNINDANGNAIP-SGELKDDIVEQIAQAKAEAGVRAQQAESNAQ 342
Db 430 PASSESTVTLPGATGCTDVLRSGLTSLPVSG-----GAVTPAPSPGGSSA 472
QY 343 ACQRYEDQHARRQEBQLQSSGIGYGLSSALIVAGGIGAGVTALHRRNQPAEQTTTTTH 402
Db 473 -----TAGPGVGSATTVAQSGATGADVL----- 495
QY 403 TVVQOQTGGIPQHKVALMP--QERRRSDRDSQGSVASTHSDSSSEVNPVAYEVGAR 460
```

```
Db 496 ----RSTSLPVGSAVPGSPGRCGATATVSSQSGPTVALSGATGTSVGP-----SGTR 547
QY 461 NLSAQHQPHEHIYDEVAADPGYGVIONFNGSGPVTGRLICTPGQGIQSTYALLANSGLR 520
Db 548 FSSSA-----IPATPG-----STTGRAAG--AGTPGVDSQQTASLPAAARPTA 588
QY 521 LGMGGLTSGGETAVSSVNAAPTPG 544
Db 589 LGPQTSAPSGET---SESRSVPG 609

RESULT 14
Q9FDA0 PRELIMINARY; PRT; 1265 AA.
AC Q9FDA0;
DT 01-MAR-2001 (T-EMBLrel. 16, Created)
DT 01-MAR-2001 (T-EMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (T-EMBLrel. 16, Last annotation update)
DE PUTATIVE OUTER MEMBRANE PROTEIN XADA.
GN XADA.
OS Xanthomonas oryzae pv. oryzae.
OC Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
OC Xanthomonas.
OX NCBI_TaxID=64187;
RN [1]
RP SEQUENCE FROM N.A.
RA Ray S.K., Rajeshwari R., Sonti R.V.;
RT "A putative outer membrane protein from Xanthomonas oryzae pv. oryzae
RT that is involved in virulence.";
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF288222; AAG01335.1; -.
SQ SEQUENCE 1265 AA; 119856 MW; 21762579B5EC70A2 CRC64;
```

```
Query Match 5.7%; Score 159.5; DB 2; Length 1265;
Best Local Similarity 23.48; Pred. No. 0.3;
Matches 136; Conservative 73; Mismatches 229; Indels 143; Gaps 29;

QY 3 IGNLGNVNGNHLIPAPPPLPSQTDGAARGTGHLISS-TGALGSRSLFSPLR-----N 55
Db 592 LGNTARALGDNSSAVGSNAVADIGATANGAGAQAALSAYTIALGSKAVASDLQAEGFN 651
QY 56 SMADSVSRDIPGLPTNPRLAAATSETCLLGGFEVLHDKGPDILNTQIGPSAFRVEQ 115
Db 652 STAGNVGSLGGSQSSGRSLSSALGYSAVSSVD-----STAVG-----VAAQ 695
QY 116 ADGTHA-AIGE--KNGLEVSVTLSPQWSSLSQSDTEGKRFVFTGGRGGHGMVTVAS 172
Db 696 ATGVSVAIGAELSKATGEESVAVGGAFSGWIPQTQASCGAAAFAGAWATADYTTAIGR 755
QY 173 D-IAEA--RTRILAKLD--PDN-----HGGROPKQVDTRSVGV-GSASGIDD-----GV 215
Db 756 DSYADGVNATAVGQSADALADNTLALGGSRRAKAVGASVIGVDASATGINSTGVGRQVNV 815
QY 216 VSEHTST-TNSSVSRDPKFWVSVCAIAAGLAGLAAATGIAQALALTPEDDPTTDDQA 274
Db 816 IGENAVSYGVSYRQS-----AVNGVALG-ANAGATG-ADSVALL---GSGSSTTDADTV 865
QY 275 ANAESATKDLQTOEAFKNPENQKVN-----DANGNAIPSGELXDDIVEQIAQAKAE 328
Db 866 SVSGNGRGGPATRI-----VNVGAGAVASASTDAINGGOLFESL----- 906
QY 329 GEVARQQAQAVESNAQAQRYEDQHARRQEBQLQSSGIGYGLSSALIVAGGIGAVVTALH- 387
Db 907 -----SNA-----ASPLGGG---AALGAGQVFVAPTYLIQG 934
QY 388 ---RRNQPAEQTTTTTHTVVQOQTGGIPQHKVALMPQERRRSDRDSQGSVASTHSDS 445
Db 935 ASYNNVGAALTALDSKVTELDARSGGTPANTAA-----RTASLRATATVPMAAATVAVSAV 988
QY 446 SSEV-----VNPVAYEVGGARNLSAQHQPHEHIYDEVAADPGYGVIONFNGSGPVTGRLICT 501
```

Db 989 SSVASTAIDATAGVGT-----PTAAVVGSTTPAAISTVVGTAANNVTGTAIG- 1039

QY 502 PCQGTQSTYALLANGGL-----RLCMGGLTS-CGTAVSSV 537

Db 1040 -----GSAYAHGANDTAIGSNARVNADGSTAVGANTQIAAV 1075

RESULT 15

Q9NHX4 PRELIMINARY; PRT: 1323 AA.

AC Q9NHX4;

DT 01-OCT-2000 (TRENBLrel. 15, Created)

DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)

DT 01-MAR-2001 (TRENBLrel. 16, Last annotation update)

DE ALHAMBRA.

GN ALHAMBRA OR CGL070.

OS Drosophila melanogaster (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephydroidea; Drosophilidae; Drosophila.

OX NCBI_taxid=7227;

RN [1]

RP SEQUENCE FROM N.A.

RA Perrin L., Dura J.M.;

RT "Alhambra, a Drosophila homolog of mammalian AF-10 and AF-17.";

RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF217960; AAF2595.1; -.

DR FlyBase; FBgn0037471; Alhambra.

DR InterPro; IPR001965; -.

DR SMART; SM00249; PHD; 1.

SQ SEQUENCE 1323 AA; 132883 MW; D53C0C8AF392F9A6 CRC64;

Query Match 5.6%; Score 156.5; DB 5; Length 1323;

Best Local Similarity 20.2%; Pred. NO. 0.47;

Matches 135; Conservative 85; Mismatches 270; Indels 177; Gaps 25;

QY 3 ICNLGNVNGNHLIPAPLPSPQTGAARGGTGHLIS-----STGALGSRSLFSP-LRNS 56

Db 322 IGINISLNN-----LPGGSSSTSSAGNVPVGGGIIISASSGGATQSTSSQSTAPGTTKS 378

QY 57 MADSVDSRDIPGLTPNPSRLAAATSETCLLGGFEVLHDKGLDILNTQIGPSAFRVEQOA 116

Db 379 SASSSSSSSSSKYKXKSLKSTSSKDKG-----KDSSTNSANNFTNSSASSTSSNS 433

QY 117 DQTHAIGKNGLEVS-----VTLSPQEWSSLOSIDPE-GKNRFVFTGGRGGG-HPM 167

Db 434 SSTR-----EKSSKLSKNKDSNQVPSATSSLTSTSSINTQPSSTSTATAGSGGTGTHVS 489

QY 168 VTVASDIAEARTRIILAKLPDNDHGGRPQKDVTRSVGVCSASGIDGV----- 215

Db 490 SAAAGLNSAPS-----TTNEHNSNHAHTLTSTNGTGAAGKLOSvNSLNSSSGFGS 542

QY 216 ----VSETHSTNTNSV-----RSDPKFWVSVGAIAAGLAGLAA--TGIAQALALTP 262

Db 543 DLRSVSTSSSTVNDSTGGFGSNSNSERENLSCAGSSASNMPGTAPGTGGVSSSAAT-- 600

QY 263 PDPPTTDPQAANAESAATKDLQEAFFKNPENOKVNIID---ANGNAIPSGELXDDIVE 319

Db 601 --NLSTNKGSSSSSTANSUTSTSSGSSSSSSKKRKADSAKSTSTSTSGSALEDNNS 658

QY 320 QIAQ-QAKEAGEVARQAVESNAQAQRYEDQHARRQEL-----QLSSGIGYGLSSA 371

Db 659 LISRYDIKDV-----QVALTPLTDFEKEIEKSKRQTELSPPHTQTSATAEVNAPLASS 713

QY 372 LIVAGGIGAGVTTALHRRNPAPAEQTTT-----TTHTVWQQQTG 410

Db 714 TSASTAVTASATAA---SAPPATGTTTLGSSISGNAGTSSGNSGVSATGGAQAVGSG 770

QY 411 GIPOHKVALMPQERRRFRDSDSGSVASTHWS-----SSSEVNVNPIYAE 455

Db 771 GYPKTESSKSSGTASAGSGSGSSNTSSPKHGSNIKIDISSSSNQQAQASTASSSAPSLYVS 830

QY 456 V-----GGARNSLSAQHOPEEHYDEVAADP----- 480

Db 831 VPLSTANVPGINLPSTSSSTTSSSHSASSSSGAQSOHQOQLSNALVGPMPGGSAGAF 890

QY 481 -----GYSVIONFSG-SGPVTGRLIGTPGOGIOSTYALLANSGLRLGMGG--LTSG 529

Db 891 HGGTTSAGSSSVIQQQSGKSSPALGTLVS-----GNSGGSIISASGFPPLPSG 937

QY 530 GETAVSS 536

Db 938 NLTAATTT 944

Search completed: September 27, 2001, 14:27:03

Job time: 314 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 27, 2001, 14:22:24 ; Search time 31.79 Seconds
(without alignments)
1066.021 Million cell updates/sec

Title: US-09-189-415A-4
Perfect score: 2851
Sequence: 1 MFIGNLGHNPVNNSTPPAP.....SNSAVNTSNPPAPGSHRFV 559

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5
Searched: 412676 seqs, 60623988 residues

Total number of hits satisfying chosen parameters: 412676

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_0601.*
1: /SIDSL/gcgdata/geneseq/geneseq/AA1980.DAT.*
2: /SIDSL/gcgdata/geneseq/geneseq/AA1981.DAT.*
3: /SIDSL/gcgdata/geneseq/geneseq/AA1982.DAT.*
4: /SIDSL/gcgdata/geneseq/geneseq/AA1983.DAT.*
5: /SIDSL/gcgdata/geneseq/geneseq/AA1984.DAT.*
6: /SIDSL/gcgdata/geneseq/geneseq/AA1985.DAT.*
7: /SIDSL/gcgdata/geneseq/geneseq/AA1986.DAT.*
8: /SIDSL/gcgdata/geneseq/geneseq/AA1987.DAT.*
9: /SIDSL/gcgdata/geneseq/geneseq/AA1988.DAT.*
10: /SIDSL/gcgdata/geneseq/geneseq/AA1989.DAT.*
11: /SIDSL/gcgdata/geneseq/geneseq/AA1990.DAT.*
12: /SIDSL/gcgdata/geneseq/geneseq/AA1991.DAT.*
13: /SIDSL/gcgdata/geneseq/geneseq/AA1992.DAT.*
14: /SIDSL/gcgdata/geneseq/geneseq/AA1993.DAT.*
15: /SIDSL/gcgdata/geneseq/geneseq/AA1994.DAT.*
16: /SIDSL/gcgdata/geneseq/geneseq/AA1995.DAT.*
17: /SIDSL/gcgdata/geneseq/geneseq/AA1996.DAT.*
18: /SIDSL/gcgdata/geneseq/geneseq/AA1997.DAT.*
19: /SIDSL/gcgdata/geneseq/geneseq/AA1998.DAT.*
20: /SIDSL/gcgdata/geneseq/geneseq/AA1999.DAT.*
21: /SIDSL/gcgdata/geneseq/geneseq/AA2000.DAT.*
22: /SIDSL/gcgdata/geneseq/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2851	100.0	559	20	AAV06221
2	1498	52.5	549	20	AAV06220
3	406	14.2	107	21	AAV06221
4	148.5	5.2	596	21	AAV99408
5	148.5	5.2	596	22	AAV99408
6	148.5	5.2	596	22	AAV99408
7	141.5	5.0	1959	12	AAV10562
8	141.5	5.0	1962	12	AAV10560
9	141.5	5.0	1962	12	AAV10561
10	141.5	5.0	1962	12	AAV10557
11	141.5	5.0	1962	12	AAV10558

12	141.5	5.0	1962	12	AAV10559	Mutant protease (A
13	141.5	5.0	1962	12	AAV10563	Mutant protease (K
14	141.5	5.0	1968	12	AAV10941	Mutant protease (d
15	141.5	5.0	1974	12	AAV10940	Mutant protease (d
16	140.5	4.9	1026	21	AAV83025	Staufen protein of
17	134.5	4.7	1638	20	AAV00138	Enterococcus faeca
18	134.5	4.7	1638	20	AAV00140	Enterococcus faeca
19	134.5	4.7	1638	20	AAV00142	Enterococcus faeca
20	134	4.7	2514	21	AAV75097	Neisseria meningit
21	133.5	4.7	2870	21	AAV95559	Caenorhabditis ele
22	133.5	4.7	3178	21	AAV95556	Caenorhabditis ele
23	133	4.7	288	20	AAV29081	T. gondii immunoge
24	133	4.7	288	20	AAV29082	T. gondii immunoge
25	133	4.7	2314	22	AAV69136	M. catarrhalis les
26	132.5	4.6	461	12	AAV14530	Usp45 protein. La
27	132.5	4.6	550	20	AAV35195	Amino acid sequenc
28	129	4.5	1048	18	AAV27277	Human cytomegalovi
29	127.5	4.5	558	20	AAV00143	Enterococcus faeca
30	125.5	4.4	461	12	AAV14150	MSP encoded by pUC
31	124.5	4.4	1529	14	AAV41732	High molecular wei
32	124.5	4.4	2048	21	AAV75096	Neisseria gonorrhe
33	124	4.3	359	19	AAV81703	M. tuberculosis im
34	124	4.3	359	19	AAV64336	Mycobacterium tube
35	124	4.3	359	20	AAV39133	M. tuberculosis an
36	124	4.3	359	20	AAV38990	M. tuberculosis re
37	123.5	4.3	1601	18	AAV30292	Non-typeable Haemo
38	123	4.3	30	20	AAV06213	EPEC E. coli trans
39	123	4.3	1098	17	AAV99392	Haemophilus adhesi
40	123	4.3	2035	15	AAV57141	Host cell factor p
41	122.5	4.3	496	20	AAV37175	Amino acid sequenc
42	121.5	4.3	456	21	AAV67238	Amino acid sequenc
43	121.5	4.3	1992	17	AAV04505	Moraxella 200 kDa
44	121.5	4.3	1992	22	AAV69133	M. catarrhalis str
45	121.5	4.3	1992	22	AAV69137	M. catarrhalis M56

ALIGNMENTS

RESULT 1
ID AAY06221 standard; Protein; 559 AA.
XX AC AAY06221;
XX AC AAY06221;
DT 16-AUG-1999 (first entry)
XX EPEC E. coli translocated intimin receptor (Tir).
DE Tir: translocated intimin receptor; Hp90; enterohaemorrhagic;
KW EHEC; infection; diagnosis; vaccine.
XX Escherichia coli.
XX Key Location/Qualifiers
FH Misc-difference 453
FT /note= "encoded by codon of 1 apparent nucleotide,
FT causing frameshift in the DNA sequence"

WO9924576-A1.
20-MAY-1999.

10-NOV-1998; 98WO-CA01042.
12-NOV-1997; 97US-0065130.
(UVR-) UNIV BRITISH COLUMBIA.
Devinney R, Finlay BB, Kenny B, Stein M;
WPI; 1999-337712/28.
N-PSDB; AAX58859.


```
Query Match          52.5%; Score 1498; DB 20; Length 549;
Best Local Similarity 54.8%; Pred. No. 6.6e-108;
Matches 323; Conservative 64; Mismatches 132; Indels 70; Gaps 12;

QY 1 MPIGNLGNPNVNSIPPAPLPSTQDGA--GGRQLINSTGPLGSRALFTPVNRSMADS 58
DB 1 mpignlgnvngnhlppapplpstqdaargtgthllsstgalgsrslfplrnsmads 60

QY 59 GDNASDVPGLVPNPMRLAA--SETLNDGEVLHDHGLPLTLNRQIGSSVFRVETQDEG 116
DB 61 vdsr--dipgiptnpsrlaatsetcllgfvevlhdkpdlintlqigsaftrvevqdg 118

QY 117 KHIAVGQRNGVETSVLSDQEVARLQSDPECKGVFTTGRGGAGHAMVTVASDITEAR 176
DB 119 thaaigeKnglevsvtlspqewsslqsldeqknrfvrtggrgshgmvtvasdiaear 178

QY 177 QRILELPEKGTG-----ESKGAGESGVGELRESNCAENTTTQTSTSTSLRS 227
DB 179 trilakldpnhggrrpkdvdtrsvvgasgi-----ddgv--vsethtstnsvrs 230

QY 228 DPKLWALGTVATGLIGLAAGIVQALALTPEPDSPTTTPDPAASATETATRDQLTKEA 287
DB 231 dpkfwsvgaiaaglaglaatgiaqalaltpepddptttddpqaanaaesatkdlqtgea 290

QY 288 FQNDPNQKVNIDELGNATPSGVKDDVANTIEEQAKAGEEAKQAIENNAQAOKKYDEQ 347
DB 291 fknpenkvndangnalpsgeikddiveiaqakeagevarqgvesnaqagryedq 350

QY 348 QAKROELKVSAGYGLSGALILGGGIVAVTAALHRKNQPVQETTTTTTTTTTSART 407
DB 351 harrqeelqlesgiyglssalivaggiagvtalhrnpaetttttt-----ht 403

QY 408 VENKPAANTPAQNVDTPCSEDTEMRSSMASTSTTFDSS-----IGGPCRIR 458
DB 404 vvqqgtgglpqhkvampqerrfrfdrdsqgsvasthwsdsssevvnpyaevvg---ar 460

QY 459 MMLMHRCMIRCRLLILRLFRWIGIISVVYSTIQHPP-----RDTTDNG---ARLLG 510
DB 461 nsishq-----peehlydevaadpyysvignfsg99pvtgrlig 500

QY 511 NPSAGIQSTYARLALSGRLDMGLTGGSNAVNTSNPPAPGSHRFV 559
DB 501 tpgqgltstyallansgglrlgmgltsgetavsvnaaptgppvrfv 549

RESULT 3
AAB20576
ID AAB20576 standard; protein; 107 AA.
XX
AC AAB20576;
XX
DT 08-DEC-2000 (first entry)
XX
DE Intimin C-terminal Tir binding domain amino acid sequence.
KW Intimin; Tir binding domain; inhibitor; intimin adhesion; screening;
KW Tir-independent eukaryotic cell binding activity; bacterial infection;
KW diarrhoea; antibacterial.
XX
OS Unidentified.
XX
PN WO200045173-A1.
XX
PD 03-AUG-2000.
XX
PF 31-JAN-2000; 2000WO-GB00254.
XX
PR 29-JAN-1999; 99GB-0001897.
XX
PA (IMCO-) IMPERIAL COLLEGE INNOVATIONS LTD.
XX
PI Frankel GM, Matthews SJ, Hale CB, Dougan G;
```

```
XX WPI; 2000-499357/44.
XX Screening for inhibitors of intimin binding to eukaryotic cells, for
PT use in diagnosing, preventing and treating bacterial infections,
PT especially Escherichia coli O157:H7
PS Claim 8; Page 76; 96pp; English.
XX
CC The present invention describes a method of screening for an inhibitor
CC of intimin binding to eukaryotic cells. The method comprises exposing an
CC intimin polypeptide having a Tir-independent cell binding activity to
CC test agents, and obtaining an inhibitor based on its ability to bind the
CC polypeptide. The inhibitors are used in the prevention, treatment and/or
CC diagnosis of bacterial infections, preferably by enteropathic and/or
CC enterohaemorrhagic Escherichia coli, Shiga toxinigenic E. coli, Hafnia
CC alvei or Citrobacter freundii, or especially E. coli O157:H7. The
CC infections cause a histopathological effect known as attachment and
CC effacement on intestinal epithelial cells. The inhibitors can be used
CC to produce food supplements or additives, especially where the food is
CC a milk substitute. The method can be used to sort cells based on their
CC ability to bind to a Tir independent cell binding domain of an intimin
CC polypeptide. Polypeptides having Tir-independent intimin binding
CC activity can be used to produce a vaccine against a bacterial disease.
CC The present sequence represents a specifically claimed intimin C-terminal
CC Tir binding domain amino acid sequence, for use in the method of the
CC present invention.
XX Sequence 107 AA;
SQ
Query Match          14.2%; Score 406; DB 21; Length 107;
Best Local Similarity 72.0%; Pred. No. 2.9e-24;
Matches 77; Conservative 18; Mismatches 12; Indels 0; Gaps 0;

QY 252 QALALTPEDSPPTTTPDPAASATETATRDQLTKEAFQNPDKVNIDELGNAIPSGVLK 311
DB 1 qalaltpepddpttdpqaanaaesatkdlqtgeafknpenkvndangnaipsgelk 60

QY 312 DDVANTIEEQAKAGEEAKQAIENNAQAOKKYDQQAQRQELKVS 358
DB 61 ddiveiaqakeagevarqgvesnaqagryedqgharrqeelqls 107

RESULT 4
AAY99408
ID AAY99408 standard; Protein; 596 AA.
XX
AC AAY99408;
XX
DT 08-AUG-2000 (first entry)
XX
DE Human PRO1342 (UNQ697) amino acid sequence SEQ ID NO:243.
KW Human; PRO polypeptide; membrane bound protein; receptor; diagnosis;
KW transmembrane; secretion; immunoadhesion; pharmaceutical; screening.
XX
OS Homo sapiens.
XX
PN WO200012708-A2.
XX
PD 09-MAR-2000.
XX
PF 01-SEP-1999; 99WO-US20111.
XX
PR 01-SEP-1998; 98US-0098716.
PR 01-SEP-1998; 98US-0098749.
PR 01-SEP-1998; 98US-0098750.
PR 02-SEP-1998; 98US-0098803.
PR 02-SEP-1998; 98US-0098821.
PR 02-SEP-1998; 98US-0098843.
PR 09-SEP-1998; 98US-0099536.
PR 09-SEP-1998; 98US-0099596.
```

PR	09-SEP-1998;	98US-0099598.	PR	22-OCT-1998;	98US-0105266.
PR	09-SEP-1998;	98US-0099602.	PR	26-OCT-1998;	98US-0105693.
PR	09-SEP-1998;	98US-0099642.	PR	26-OCT-1998;	98US-0105694.
PR	10-SEP-1998;	98US-0099741.	PR	27-OCT-1998;	98US-0105807.
PR	10-SEP-1998;	98US-0099754.	PR	27-OCT-1998;	98US-0105881.
PR	10-SEP-1998;	98US-0099763.	PR	27-OCT-1998;	98US-0105882.
PR	10-SEP-1998;	98US-0099792.	PR	27-OCT-1998;	98US-0106062.
PR	10-SEP-1998;	98US-0099808.	PR	28-OCT-1998;	98US-0106023.
PR	10-SEP-1998;	98US-0099812.	PR	28-OCT-1998;	98US-0106029.
PR	10-SEP-1998;	98US-0099815.	PR	28-OCT-1998;	98US-0106030.
PR	15-SEP-1998;	98US-0099816.	PR	28-OCT-1998;	98US-0106032.
PR	15-SEP-1998;	98US-0100385.	PR	28-OCT-1998;	98US-0106033.
PR	15-SEP-1998;	98US-0100388.	PR	28-OCT-1998;	98US-0106178.
PR	15-SEP-1998;	98US-0100390.	PR	28-OCT-1998;	98US-0106248.
PR	16-SEP-1998;	98US-0100584.	PR	29-OCT-1998;	98US-0106384.
PR	16-SEP-1998;	98US-0100627.	PR	29-OCT-1998;	98US-0106500.
PR	16-SEP-1998;	98US-0100661.	PR	30-OCT-1998;	98US-0106464.
PR	16-SEP-1998;	98US-0100662.	PR	30-OCT-1998;	98US-0106856.
PR	16-SEP-1998;	98US-0100664.	PR	03-NOV-1998;	98US-0106902.
PR	17-SEP-1998;	98US-0100683.	PR	03-NOV-1998;	98US-0106905.
PR	17-SEP-1998;	98US-0100684.	PR	03-NOV-1998;	98US-0106919.
PR	17-SEP-1998;	98US-0100710.	PR	03-NOV-1998;	98US-0106932.
PR	17-SEP-1998;	98US-0100711.	PR	03-NOV-1998;	98US-0106934.
PR	17-SEP-1998;	98US-0100919.	PR	10-NOV-1998;	98US-0107783.
PR	17-SEP-1998;	98US-0100930.	PR	17-NOV-1998;	98US-0108775.
PR	18-SEP-1998;	98US-0100848.	PR	17-NOV-1998;	98US-0108779.
PR	18-SEP-1998;	98US-0100849.	PR	17-NOV-1998;	98US-0108787.
PR	18-SEP-1998;	98US-0101014.	PR	17-NOV-1998;	98US-0108788.
PR	18-SEP-1998;	98US-0101068.	PR	17-NOV-1998;	98US-0108801.
PR	18-SEP-1998;	98US-0101071.	PR	17-NOV-1998;	98US-0108802.
PR	22-SEP-1998;	98US-0101279.	PR	17-NOV-1998;	98US-0108806.
PR	23-SEP-1998;	98US-0101471.	PR	17-NOV-1998;	98US-0108807.
PR	23-SEP-1998;	98US-0101472.	PR	17-NOV-1998;	98US-0108867.
PR	23-SEP-1998;	98US-0101473.	PR	17-NOV-1998;	98US-0108925.
PR	23-SEP-1998;	98US-0101475.	PR	18-NOV-1998;	98US-0108848.
PR	23-SEP-1998;	98US-0101476.	PR	18-NOV-1998;	98US-0108849.
PR	23-SEP-1998;	98US-0101477.	PR	18-NOV-1998;	98US-0108850.
PR	23-SEP-1998;	98US-0101479.	PR	18-NOV-1998;	98US-0108851.
PR	24-SEP-1998;	98US-0101738.	PR	18-NOV-1998;	98US-0108852.
PR	24-SEP-1998;	98US-0101741.	PR	18-NOV-1998;	98US-0108858.
PR	24-SEP-1998;	98US-0101743.	PR	18-NOV-1998;	98US-0108904.
PR	24-SEP-1998;	98US-0101915.	XX		
PR	24-SEP-1998;	98US-0101916.	XX		
PR	29-SEP-1998;	98US-0102207.	XX		
PR	29-SEP-1998;	98US-0102240.	XX		
PR	29-SEP-1998;	98US-0102307.	XX		
PR	29-SEP-1998;	98US-0102330.	XX		
PR	29-SEP-1998;	98US-0102331.	XX		
PR	29-SEP-1998;	98US-0102484.	XX		
PR	30-SEP-1998;	98US-0102487.	XX		
PR	30-SEP-1998;	98US-0102570.	XX		
PR	30-SEP-1998;	98US-0102571.	XX		
PR	01-OCT-1998;	98US-0102684.	XX		
PR	01-OCT-1998;	98US-0102687.	XX		
PR	02-OCT-1998;	98US-0102965.	XX		
PR	06-OCT-1998;	98US-0103258.	XX		
PR	06-OCT-1998;	98US-0103449.	XX		
PR	07-OCT-1998;	98US-0103314.	XX		
PR	07-OCT-1998;	98US-0103315.	XX		
PR	07-OCT-1998;	98US-0103328.	XX		
PR	07-OCT-1998;	98US-0103395.	XX		
PR	07-OCT-1998;	98US-0103396.	XX		
PR	07-OCT-1998;	98US-0103401.	XX		
PR	08-OCT-1998;	98US-0103633.	XX		
PR	08-OCT-1998;	98US-0103678.	XX		
PR	08-OCT-1998;	98US-0103679.	XX		
PR	08-OCT-1998;	98US-0103711.	XX		
PR	14-OCT-1998;	98US-0104257.	XX		
PR	20-OCT-1998;	98US-0104987.	XX		
PR	20-OCT-1998;	98US-0105000.	XX		
PR	20-OCT-1998;	98US-0105002.	XX		
PR	21-OCT-1998;	98US-0105104.	XX		
PR	22-OCT-1998;	98US-0105169.	XX		
					Sequence 596 AA;
					Query Match 5.2%; Score 148.5; DB 21; Length 596;
					Best Local Similarity 17.7%; Pred No. 0.0029;
					Matches 102; Conservative 84; Mismatches 255; Indels 135; Gaps 19;
					4 GNLGHPNVNNSIPPAPLPISQTDGAGGQLINSTGPIGSRALFTPVNSMADSGDNRA 63

(GETH) GENENTECH INC.

Baker K, Goddard A, Gurney AL, Smith V, Watanabe CK, Wood WI;

WPI; 2000-237871/20.

N-PSDB; AAA37090.

New mammalian DNA sequences encoding transmembrane, receptor or secreted PRO polypeptides, useful for screening of potential peptide or small molecule inhibitors of the relevant receptor/ligand interactions

Claim 12; Fig 138; 773pp; English.

AAA37022 to AAA37144 encode the new isolated human transmembrane, receptor or secreted PRO polypeptides given in AA99340 to AA99462. The transmembrane and receptor PRO proteins can be used for screening of potential peptide or small molecule inhibitors of the relevant receptor/ligand interactions. The polypeptides and nucleotide sequences encoding then have various industrial applications, including uses as pharmaceutical and diagnostic agents. AAA37145 to AAA37330 represent PCR primers and hybridisation probes used in the isolation of the PRO polypeptides from the present invention.

```
Db 13 gillhleatns-----netsantgssvissg-----astatnsg 49
QY 64 SDVPGLPVNPMLAASETLNDGFVLDHDPGLDPLNRQIGSSV-FRVETQEDGKHIAVG 122
Db 50 ssvtsgvstatnsgsvtsn-gvsiv-tnsefhtstgstatnsefstassgisiatn 107
QY 123 QRNGVETS--VVLSDQEVARLQSIDPEGKDFVFTGGRGGAGHAMVTVASDITEARQRL 180
Db 108 sessttsgastatnse-----sstpsgastvtnsgsvtsgastatnseestvsrras 163
QY 181 ELLEPKGTGSGAGESKGVGELRESNAGENTTQTSTSSLSRSDPKLWALGTAVT 240
Db 164 tatnsestlsgastatn-sdssttsgastatnsesttsg-----astat 211
QY 241 GLIGLAATGIVQALALTPEDSPPTTDPDAAASATETATRDQLTKAEAFONPNQKNIDE 300
Db 212 n-----sesstvsrrastatnsestt---sgastatnsesttngagatnsestts 264
QY 301 LG-----NAIPSGVLKDDVAVANIEEQAKAAGEAKQQAENNAQAOKKYDQOAKR 351
Db 265 sgastatnsestvsrga---statnsesttsgast-----atn 302
QY 352 QEELKVSSGAGYGL-SGALILGGGIGVAVTAALHRKNQPVQETTTTTTTTSARTVEN 410
Db 303 sessttsgastatnsesttsgastatnsesttsgastatnsesttsgantatn 362
QY 411 KPANTPAQGNVDTGSEDTHMESRRSSMASTSTFTDSSIGGPCRIRMLMLKHCMIIR 470
Db 363 sessttsgastatnsesttsgastatnsesttsgastatnsesttsgantatn 401
QY 471 CRLLILRLFRWIGIQISVVYSTIOHPPRDTHDNGARLLGNPSAGIQSTYARLA----- 524
Db 402 -----vstatnsesttsgastatnsesttsgastatnsesttsgastatnsest 441
QY 525 -LSGGL----RHDGGLTGGSNSAVNTSNPPAPGS 555
Db 442 tvssgistvtneesttsgantatnsgsvtsags 477

RESULT 5
AAB87575
ID AAB87575 standard; Protein; 596 AA.
AC AAB87575;
XX
XX 15-MAY-2001 (first entry)
DE Human PRO1342.
KW Human; PRO protein; mapping.
OS Homo sapiens.
XX
XX WO200116318-A2.
PD 08-MAR-2001.
PF 24-AUG-2000; 2000WO-US23328.
XX
XX 01-SEP-1999; 99WO-US20111.
PR 15-SEP-1999; 99WO-US21090.
PR 07-DEC-1999; 99US-0169495.
PR 09-DEC-1999; 99US-0170262.
PR 11-JAN-2000; 2000US-0175481.
PR 18-FEB-2000; 2000WO-US04341.
PR 18-FEB-2000; 2000WO-US04342.
PR 22-FEB-2000; 2000WO-US04414.
PR 01-MAR-2000; 2000WO-US05601.
PR 03-MAR-2000; 2000US-0187202.
PR 25-APR-2000; 2000US-0199397.
PR 22-MAY-2000; 2000WO-US14042.
PR 05-JUN-2000; 2000US-0209832.
```

```
XX
PA (GETH ) GENENTECH INC.
XX
PI Eaton DL, Filvaroff E, Gerritsen ME, Goddard A, Godowski PJ;
PI Grimaldi CJ, Gurney AL, Watanabe CK, Wood WI;
XX
XX WPI; 2001-183260/18.
DR N-PSDB; AAF92107.
XX
XX Eighty four nucleic acids encoding PRO polypeptides, useful in
PT molecular biology, including use as hybridization probes, and in
PT chromosome and gene mapping.
XX
PS Claim 12; Fig 100; 278pp; English.
XX
CC The present sequence is a human PRO polypeptide (secreted and
CC transmembrane). The PRO protein, and PRO agonists, PRO antagonists or
CC anti-PRO antibodies are useful for preparation of a medicament useful in
CC the treatment of a condition which is responsive to the PRO protein,
CC agonists, antagonists or anti-PRO antibodies. The PRO protein may also be
CC employed as molecular weight markers for protein electrophoresis. The PRO
CC coding sequence has applications in molecular biology, including use as
CC hybridisation probes, and in chromosome and gene mapping.
XX
SQ Sequence 596 AA;
```

Query Match 5.2%; Score 148.5; DB 22; Length 596;

Best Local Similarity 17.7%; Pred. No. 0.0029;

Matches 102; Conservative 84; Mismatches 255; Indels 135; Gaps 19;

```
QY 4 GNIGHNPVNNISIPAPLPSTODGAGRGQLINSTGPLGSRALFTPVNSMADSGDRA 63
Db 13 gillhleatns-----netsantgssvissg-----astatnsg 49
QY 64 SDVPGLPVNPMLAASETLNDGFVLDHDPGLDPLNRQIGSSV-FRVETQEDGKHIAVG 122
Db 50 ssvtsgvstatnsgsvtsn-gvsiv-tnsefhtstgstatnsefstassgisiatn 107
QY 123 QRNGVETS--VVLSDQEVARLQSIDPEGKDFVFTGGRGGAGHAMVTVASDITEARQRL 180
Db 108 sessttsgastatnse-----sstpsgastvtnsgsvtsgastatnseestvsrras 163
QY 181 ELLEPKGTGSGAGESKGVGELRESNAGENTTQTSTSSLSRSDPKLWALGTAVT 240
Db 164 tatnsestlsgastatn-sdssttsgastatnsesttsg-----astat 211
QY 241 GLIGLAATGIVQALALTPEDSPPTTDPDAAASATETATRDQLTKAEAFONPNQKNIDE 300
Db 212 n-----sesstvsrrastatnsestt---sgastatnsesttngagatnsestts 264
QY 301 LG-----NAIPSGVLKDDVAVANIEEQAKAAGEAKQQAENNAQAOKKYDQOAKR 351
Db 265 sgastatnsestvsrga---statnsesttsgast-----atn 302
QY 352 QEELKVSSGAGYGL-SGALILGGGIGVAVTAALHRKNQPVQETTTTTTTTSARTVEN 410
Db 303 sessttsgastatnsesttsgastatnsesttsgastatnsesttsgantatn 362
QY 411 KPANTPAQGNVDTGSEDTHMESRRSSMASTSTFTDSSIGGPCRIRMLMLKHCMIIR 470
Db 363 sessttsgastatnsesttsgastatnsesttsgastatnsesttsgantatn 401
QY 471 CRLLILRLFRWIGIQISVVYSTIOHPPRDTHDNGARLLGNPSAGIQSTYARLA----- 524
Db 402 -----vstatnsesttsgastatnsesttsgastatnsesttsgastatnsest 441
QY 525 -LSGGL----RHDGGLTGGSNSAVNTSNPPAPGS 555
Db 442 tvssgistvtneesttsgantatnsgsvtsags 477
```

RESULT 6

CC The mutant protease K13/G/K13D having new cleavage of casein

Db 1840 tkpattsttdtdtrngltsdskgggqtpapapgdickdkgdegspssggnip 1899

QY 439 ----ASTSSTFFDTSSIGG 453
ID AAR10559 standard; Protein; 1962 AA.
Db 1900 tnpattsttdtdtrng 1918

RESULT 12
AAR10559
XX AAR10559 standard; Protein; 1962 AA.
AC AAR10559;
XX AAR10559;
DT 15-APR-1991 (first entry)

DE Mutant protease (A137G/K138L/T139A).
XX
KW Mutant protease gene; fermentation; foodstuff; flavouring;
KW lactic acid bacteria.
XX
OS Lactococcus lactis SK11.

XX
FH Key Location/Qualifiers
FT Peptide 1..187
FT Protein 188..1962
FT /label= signal_peptide
FT /label= mature_protein

XX EP411715-A.
XX
XX 06-FEB-1991.
XX
XX 02-AUG-1990; 90EP-0202113.
XX
XX 04-AUG-1989; 89NL-0002010.
XX
XX (NEZU-) NED INST ZUIVELONDE.

XX Vos PAJ, Sieren RJ, De Vos WM, Kok J, Venema G, Haandrikman AJ;
PI WPI; 1991-038622/06.
DR N-PSDB; AAQ10413.
XX
XX Mutant protease gene(s) and protease(s) - derived from type I and
PT III protease genes from lactococcal strains, used in fermentation
PT foodstuffs and flavourings

XX
XX Disclosure; Fig 1(1-7)+5(b); 29 pp; English.
XX
XX The mutant protease having new cleavage specificities is
CC obtained by replacing three amino acids.
CC This mutant may then be used to prepare hybrid proteases,
CC the fusion being between a type I and a type III protease of
CC L.lactis Wg2 and SK11.

CC The product has modified properties, e.g. thermostability,
CC alkaline/acid pH stability, oxidative stability, autolysis etc.,
CC compared to the parent protease(s). The proteases can be used for
CC preparing products (butter cheese, human and animal foodstuffs)
CC prepared with the aid of lactic acid bacteria.
CC See also AAQ10411-17 and AAQ10870-71.

XX
XX Sequence 1962 AA;
XX
XX Query Match 5.0%; Score 141.5; DB 12; Length 1962;
XX Best Local Similarity 22.8%; Pred. No. 0.051;
XX Matches 114; Conservative 61; Mismatches 211; Indels 113; Gaps 20;

QY 36 INSTGPIGSRALFTVPR-----NSMADSGDNRASDPVGLPVNPMR-LAASEITLND 85
Db 1452 insgkp-gmaidqpvkllgknvltavtdsedntttnitvyyepkktlaaptvtps- 1509
QY 86 GFEVLHDHGPDLTLNRIGSSVFRVETQEDGKHIAVGQRNGVETSVLSQDEYARLQSID 145

Db 1510 -----ttepactvtltanaaa-----tgetvgsadgktyqdvdp-----aagvtit 1551

QY 146 PEGKDKFVFTGGGAGHAMVTVASDI-----TEARQRIELLEPKGTGESKAGES 197
Db 1552 angtfkfstldlygnespavdyvtnikaddpaqiqaakqeltnliasaaktlisagkydd 1611

QY 198 KGVGELRESNGAENTETQTSTSTSSLR-SDPKLWALGTAVTGLIGLAATGIIVQALAL 256
Db 1612 atttalaatqkaq-caldqtnasvdsiltganrdlqtainqlaaklpadkkttsllnqlqs 1670

QY 257 TPE----PDSPTTIDPDAASATETATPDOLTKEA---FONPDNQKNIDELGNAIPSGVL 310
Db 1671 vkdalgtldlgnqtdp--stgkftfaalddlvagaqagtdtdqlgtiak-----i 1719

QY 311 KDDVWANIIEQAKAAGE-----EAKQQAII-----ENNAOAKKYDEQOAKR 351
Db 1720 ldevlaklaegikaatpaevgnakdaatgktyadiadtltsgasadasdklahlqaiq 1779

QY 352 QBELKVSS-----GAGYGLSGALILGGIGVAVTA-----ALHRKNQP 389
Db 1780 slktkvaavaeaaakvgkgdgttgsdkgggqtpapapgdtkdkgdegspssggnip 1839

QY 390 VEQTTFITTTTTSAR-----TVENKPNANTPAQGNVD---TPGSEDTMESRRSSM- 438
Db 1840 tkpattsttdtdtrngltsdskgggqtpapapgdickdkgdegspssggnip 1899

QY 439 ----ASTSSTFFDTSSIGG 453
Db 1900 tnpattsttdtdtrng 1918

RESULT 13
AAR10563
ID AAR10563 standard; Protein; 1962 AA.
XX
XX AAR10563;
AC AAR10563;
XX
XX 15-APR-1991 (first entry)
XX
XX Mutant protease (K748T).
XX
XX Mutant protease gene; fermentation; foodstuff; flavouring;
KW lactic acid bacteria.
KW
OS Lactococcus lactis SK11.

XX
XX Key Location/Qualifiers
FT Peptide 1..187
FT Protein 188..1962
FT /label= signal_peptide
FT /label= mature_protein
XX
XX EP411715-A.
XX
XX 06-FEB-1991.
XX
XX 02-AUG-1990; 90EP-0202113.
XX
XX 04-AUG-1989; 89NL-0002010.
XX
XX (NEZU-) NED INST ZUIVELONDE.

XX Vos PAJ, Sieren RJ, De Vos WM, Kok J, Venema G, Haandrikman AJ;
PI WPI; 1991-038622/06.
DR N-PSDB; AAQ10417.
XX
XX Mutant protease gene(s) and protease(s) - derived from type I and
PT III protease genes from lactococcal strains, used in fermentation
PT foodstuffs and flavourings
XX
XX Disclosure; Fig 1(1-7)+5(a); 29 pp; English.

XX The mutant protease K748T having new cleavage specificities is
 CC obtained by carrying out single amino acid substitutions.
 CC This mutant may then be used to prepare hybrid proteases,
 CC the fusion being between a type I and a type III protease of
 CC L.lactis Wg2 and SK11.
 CC The product has modified properties, e.g. thermostability,
 CC alkaline/acid pH stability, oxidative stability, autolysis etc.,
 CC compared to the parent protease(s). The proteases can be used for
 CC preparing products (butter cheese, human and animal foodstuffs)
 CC prepared with the aid of lactic acid bacteria.
 CC See also AAQ10411-17 and AAQ10870-71.
 XX
 SQ Sequence 1962 AA;

Query Match 5.0%; Score 141.5; DB 12; Length 1962;
 Best Local Similarity 22.8%; Pred. No. 0.051;
 Matches 114; Conservative 61; Mismatches 211; Indels 113; Gaps 20;

QY 36 INSTGPGSRALFTVPR-----NSMADSGDNRASDPVGLPVNPMR-LAASEITLND 85
 Db 1452 insgkp-ghmaidqpvklllegknvltavtdehntttknitvyyepkktlaaptvtps- 1509
 QY 86 GFEVLHDHGFLDLNRQIGSSVFRVETQEDGKHIAVGQRNGVETSVVLSDOEYARLQSID 145
 Db 1510 -----ttepactvtltanaaa---tgetvqysadgktyqdv-----aagvtit 1551
 QY 146 PGKDKFVETGGGAGHAMVTVASDI-----TEARQRIELLEPKGTGESKAGES 197
 Db 1552 angtfkfstldlygnespavdyvntnikaddpaqlqaakqeltnliasaaktsasgkydd 1611
 QY 198 KGVGELRESNSGAENTETQSTSTSSLR-SDPKLWALGTATGLIGLAANGIVQALAL 256
 Db 1612 attalaaatqkaq-taldqtnasvdsldgandrdlqtainglaaklpadkktlsllnqlqs 1670
 QY 257 TPE---PDSPTTTDPDAAASATETATRDOLTKEA---FQNPQNKVNIDELGNAPSGVL 310
 Db 1671 vkdalgtldlgnqtdp--sgkfttaaldldlvaqaqagtqtdqqlqatlak-----i 1719
 QY 311 KDDVVANIEEQAKAGE-----EAKQQA-----ENNAQAKKYDEQQAQR 351
 Db 1720 ldeviaklaegikaatpaevgnakdaatgktywadiadltltsgqasadasdklahlqalq 1779
 QY 352 QBELKVSS-----GAGYGLSGALLILGGIGVAVTA-----ALHRKNQP 389
 Db 1780 slktkvaaveaaktvgkgdgttsdkggggtgtpapapgdgtgkdgdegspssggnip 1839
 QY 390 VEQTTTTTTTTTTTSAR-----TVENKPANNTPAQGNVD---TPGSBDTMSRRSSM- 438
 Db 1840 tkpattsttdtdtdrngqltsdtkggggtgtpapapgdgkdgdegspssggnip 1899
 QY 439 ----ASTSTFTDTSIGG 453
 Db 1900 tnpattsttdtdtdrng 1918

RESULT 14
 AAR10941
 ID AAR10941 standard; Protein; 1968 AA.
 XX
 AC AAR10941;
 XX
 DT 15-APR-1991 (first entry)
 XX
 DE Mutant protease (delta137-139/ins9).
 XX
 KW Mutant protease gene; fermentation; foodstuff; flavouring;
 OS lactic acid bacteria.
 XX
 OS Lactococcus lactis SK11.
 XX
 FH Key Location/Qualifiers

FT Peptide 1..187
 FT Protein /label= sig_peptide
 FT 188..1968
 FT /label= mat_protein
 XX
 PN EP411715-A.
 XX
 PD 06-FEB-1991.
 XX
 PF 02-AUG-1990; 90EP-0202113.
 XX
 PR 04-AUG-1989; 89NL-0002010.
 XX
 PA (NEZU-) NED INST ZUIVELONDE.
 XX
 PI Vos PAJ, Sieren RJ, De Vos WM, Kok J, Venema G, Haandrikman AJ;
 XX
 DR WPI; 1991-038622/06.
 DR N-PSDB; AAQ10871.
 XX
 PT Mutant protease gene(s) and protease(s) - derived from type I and
 PT III protease genes from lactococcal strains, used in fermentation
 PT foodstuffs and flavourings
 XX
 PS Disclosure: Fig 1(1-7)+5(b); 29pp; English.
 XX
 CC The wild-type L.lactis SK11 protease gene sequence was determined
 CC by the applicant (EP-307011).
 CC The mutant protease having new cleavage specificities is
 CC obtained by deleting three amino acids (nine bps) and inserting
 CC 9 other residues.
 CC This mutant may then be used to prepare hybrid proteases,
 CC the fusion being between a type I and a type III protease of
 CC L.lactis Wg2 and SK11.
 CC The product has modified properties, e.g. thermostability,
 CC alkaline/acid pH stability, oxidative stability, autolysis etc.,
 CC compared to the parent protease(s). The proteases can be used for
 CC preparing products (butter cheese, human and animal foodstuffs)
 CC prepared with the aid of lactic acid bacteria.
 CC See also AAQ10411-17 and AAQ10870-71.
 XX
 SQ Sequence 1968 AA;

Query Match 5.0%; Score 141.5; DB 12; Length 1968;
 Best Local Similarity 22.8%; Pred. No. 0.052;
 Matches 114; Conservative 61; Mismatches 211; Indels 113; Gaps 20;

QY 36 INSTGPGSRALFTVPR-----NSMADSGDNRASDPVGLPVNPMR-LAASEITLND 85
 Db 1458 insgkp-ghmaidqpvklllegknvltavtdehntttknitvyyepkktlaaptvtps- 1515
 QY 86 GFEVLHDHGFLDLNRQIGSSVFRVETQEDGKHIAVGQRNGVETSVVLSDOEYARLQSID 145
 Db 1516 -----ttepactvtltanaaa---tgetvqysadgktyqdv-----aagvtit 1557
 QY 146 PGKDKFVETGGGAGHAMVTVASDI-----TEARQRIELLEPKGTGESKAGES 197
 Db 1558 angtfkfstldlygnespavdyvntnikaddpaqlqaakqeltnliasaaktsasgkydd 1617
 QY 198 KGVGELRESNSGAENTETQSTSTSSLR-SDPKLWALGTATGLIGLAANGIVQALAL 256
 Db 1618 attalaaatqkaq-taldqtnasvdsldgandrdlqtainglaaklpadkktlsllnqlqs 1676
 QY 257 TPE---PDSPTTTDPDAAASATETATRDOLTKEA---FQNPQNKVNIDELGNAPSGVL 310
 Db 1677 vkdalgtldlgnqtdp--sgkfttaaldldlvaqaqagtqtdqqlqatlak-----i 1725
 QY 311 KDDVVANIEEQAKAGE-----EAKQQA-----ENNAQAKKYDEQQAQR 351
 Db 1726 ldeviaklaegikaatpaevgnakdaatgktywadiadltltsgqasadasdklahlqalq 1785
 QY 352 QBELKVSS-----GAGYGLSGALLILGGIGVAVTA-----ALHRKNQP 389

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 27, 2001, 14:21:04 ; Search time 21.2 seconds
(without alignments)
542.925 Million cell updates/sec

Title: US-09-189-415A-4

Perfect score: 2851

Sequence: 1 MP1GNLGHNPVNSIPPAP.....SNSAVNTSNPPAGSHREV 559

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 197339 seqs, 20590346 residues

Total number of hits satisfying chosen parameters: 197339

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents_AA:*

1: /cgn2_6/ptodata/2/1aa/5A_COMB.pep.*
2: /cgn2_6/ptodata/2/1aa/5B_COMB.pep.*
3: /cgn2_6/ptodata/2/1aa/6A_COMB.pep.*
4: /cgn2_6/ptodata/2/1aa/6B_COMB.pep.*
5: /cgn2_6/ptodata/2/1aa/PCRTUS_COMB.pep.*
6: /cgn2_6/ptodata/2/1aa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Length	DB ID	Description
1	125.5	4.4	461	1	US-08-186-222-2
2	125	4.4	1702	4	US-08-296-791-5
3	125	4.4	1702	5	PCT-US95-10661A-5
4	124.5	4.4	1529	2	US-08-728-470-10
5	124.5	4.4	1529	4	US-08-719-641-10
6	124.5	4.4	1600	2	US-08-617-697-10
7	123	4.3	658	1	US-08-409-995-5
8	123	4.3	658	3	US-08-685-467-5
9	123	4.3	658	4	US-08-913-942-5
10	123	4.3	1098	1	US-08-409-995-2
11	123	4.3	1098	3	US-08-685-467-2
12	123	4.3	1098	4	US-09-377-155-32
13	123	4.3	1098	4	US-08-913-942-2
14	123	4.3	1848	4	US-08-296-791-6
15	123	4.3	1848	5	PCT-US95-10661A-6
16	123	4.3	2035	1	US-08-046-585-5
17	123	4.3	2035	1	US-08-393-703-5
18	123	4.3	2035	5	PCT-US93-11721-5
19	120	4.2	675	1	US-08-386-495-10
20	120	4.2	675	5	PCT-US96-02331-10
21	120	4.2	842	5	PCT-US96-02331-15
22	119	4.2	1536	1	US-08-038-682-2
23	119	4.2	1536	1	US-08-302-832-2
24	119	4.2	1536	2	US-08-530-198-2
25	119	4.2	1536	2	US-08-469-880-2
26	119	4.2	1536	2	US-08-728-470-2
27	119	4.2	1536	2	US-08-617-697-2

28 119 4.2 1536 4 US-08-719-641-2 Sequence 2, Appli
29 119 4.2 1581 4 US-09-110-517-2 Sequence 2, Appli
30 118 4.1 1780 1 US-08-769-309A-5 Sequence 5, Appli
31 118 4.1 1780 3 US-08-994-570-5 Sequence 5, Appli
32 117.5 4.1 694 3 US-08-559-397A-31 Sequence 31, Appli
33 117.5 4.1 1026 2 US-08-614-377A-7 Sequence 7, Appli
34 117.5 4.1 1026 4 US-09-142-648B-7 Sequence 7, Appli
35 116 4.1 546 3 US-08-935-855-20 Sequence 20, Appli
36 115.5 4.1 985 5 PCT-US96-03916-6 Sequence 6, Appli
37 115.5 4.1 985 5 PCT-US96-03916-6 Sequence 66, Appli
38 115.5 4.1 1338 2 US-08-728-470-9 Sequence 9, Appli
39 115.5 4.1 1338 4 US-08-719-641-9 Sequence 9, Appli
40 115.5 4.1 1599 2 US-08-617-697-9 Sequence 9, Appli
41 115 4.0 484 1 US-08-127-499A-26 Sequence 26, Appli
42 115 4.0 484 1 US-08-482-847-26 Sequence 26, Appli
43 111.5 3.9 631 4 US-08-847-065-25 Sequence 25, Appli
44 111.5 3.9 1026 1 US-08-194-290-7 Sequence 7, Appli
45 111 3.9 2353 4 US-09-377-155-33 Sequence 33, Appli

ALIGNMENTS

RESULT 1
US-08-186-222-2
; Sequence 2, Application US/08186222
; Patent No. 5559007
; GENERAL INFORMATION:
; APPLICANT: Suri, Bruno
; APPLICANT: Schmitz, Albert
; TITLE OF INVENTION: Bacterial Vectors
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CIBA-GEIGY Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: New York
; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/186,222
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/672,205
; FILING DATE: 19-MAR-1991
; APPLICATION NUMBER: GB 9006400.7
; FILING DATE: 22-MAR-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Villamizar, JoAnn
; REGISTRATION NUMBER: 30,598
; REFERENCE/DOCKET NUMBER: 4-17994/A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (914)785-7121
; TELEFAX: (914)347-5769
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 461 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-186-222-2

Query Match 4.4%; Score 125.5; DB 1; Length 461;
Best Local Similarity 19.0%; Pred. No. 0.0051;
Matches 84; Conservative 70; Mismatches 160; Indels 127; Gaps 16;

US-08-296-791-5

RESULT 3

PC1 0055 10061A 5
: sequence 5. Application PC/TUS9510661A

APPLICANT: Washington University, et al.

NUMBER OF SEQUENCES:

; ADDRESS: Flehr, H

CITY: San Francisco

COUNTRY: United States

```

;
; COMPUTER READABLE FORM
; VECTOR TYPE: F1 CODE

```

COMPUTER: IBM PC
OPERATING SYSTEM: PC DOS 2.0

SOFTWARE: TASCHEEN
CURRENT APPLICATION DATA

FILING DATE: 16-AUG-2011

PRIOR APPLICATION DATA

FILING DATE: 25-AUG-

;
ATTORNEY/AGENT INFORM

; REGISTRATION NUMBER:

REFERENCE/DOCKET NUMBER: FP-59941/RFT
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
TELEX: 910 277299
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 1702 amino acids
TYPE: amino acid
TOPOLOGY: unknown
PCT-US95-10661A-5

Query Match 4.4%; Score 125; DB 5; Length 1702;
Best Local Similarity 17.8%; Pred. No. 0.044;
Matches 87; Conservative 62; Mismatches 180; Indels 160; Gaps 17;
QY 10 PNVNNSI-----PPAPPLPSQTDGAGRGQLINSTPLGSRALFTPVNRSMADSGDN 61
DB 1017 PSNNEELARVETVPVPPAPASETE-----TVAENSKQES--- 1053
QY 62 RASDVPGLVPMRLAASITLNDGFEVLHDHGPLDTLNRQIGSSVFRVETQEDGKHIAV 121
DB 1054 -----KTVKNEQDATETTAQNG-EVAEEAKPSVKANTQTN-----EVAQSGSETE 1099
QY 122 GQNGVETSVLSDQEARLQSIDPEGRDKFVFTGGRGAGHAMVTVASDITEARQRILE 181
DB 1100 TQTEIKETAKVEKEEKAKEKAKVEK-----DEIQEAPQMASE 1141
QY 182 LLEPKGTGSKGAGESKGVGELR-ESNSGAENTT-----ETQTSSTSSL 225
DB 1142 TSPKQAKPAKEVSTDTKVEETQVQAQPTQSTYVAABATSPNKPARETOPSEKTN-- 1199
QY 226 RSDPKLWALGVATGLIGLANTGIVQALALTPEDSPPTTDPDAAASATETATRDQLTK 285
DB 1200 -AEP-----VTPVVKQNTENTDQPTEREKTAKEKTEKQEPQVAS 1241
QY 286 EAFQPNQKV-----NIDELGNAIPSGVLKDDVVANIEEQAKAAGEAKQOATE 335
DB 1242 QASPKQEQSETVQPAVLESENVPTVNN-----EEVQAQLOQTQTSAT-VSTKQPAPE 1293
QY 336 NNAQAQKYYDEQAQRQEELKYSVGAGYGLSGALILGGGIGVAVTAALHRKNQVQETTT 395
DB 1294 NS-----INTGSATAITAEKSKDPQTETAA 1320
QY 396 TTTTITTTTSARTV-ENKANNTPA-----QGNVDTPGSEDTMESRSSMASTSTFPD 447
DB 1321 STEDASQHKANTVADNSVANNSESSEPKSRRRRSISQP--QETSABETTAATSTDETTIAD 1378
QY 448 TSSIGGPCR 456
DB 1379 NSKRKPNR 1387

RESULT 4
US-08-728-470-10
Sequence 10, Application US/08728470
Patent No. 5928651
GENERAL INFORMATION:
APPLICANT: Barenkamp, Stephen J
TITLE OF INVENTION: High Molecular Weight Surface Proteins
TITLE OF INVENTION: of No. 5928651-Typeable Haemophilus
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Shoemaker and Mattare, Ltd.
STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202-0286
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/728,470
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/302,832
FILING DATE: 16-MAR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US PCT/US93/02166
FILING DATE: 16-MAR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9205704.1
FILING DATE: 16-MAR-1992
ATTORNEY/AGENT INFORMATION:
NAME: Berkstresser, Jerry W
REGISTRATION NUMBER: 22,651
REFERENCE/DOCKET NUMBER: 1038-633
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 415-0810
TELEFAX: (703) 415-0813
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 1529 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-728-470-10

Query Match 4.4%; Score 124.5; DB 2; Length 1529;
Best Local Similarity 20.2%; Pred. No. 0.041;
Matches 123; Conservative 76; Mismatches 233; Indels 177; Gaps 28;
QY 26 TDGAG-----RGOLINSTPLGSRALFTPVNRSMADS-----GDNRASDVPGLPVPNMR 75
DB 882 TYNASGTQTIINGNTNEKGLN-----TKNIKADAEIQIGNISOKEGNLTISSOK 934
QY 76 L-AASEITLNDGFE-VLHDHGPLDTLNRQIGSSVFRVE-----TQEDGKHIA 120
DB 935 VNITNQITIKAGVEGSRSDSEANANLTQTKELAGDLNISGENKAEITAKNGSDLT 994
QY 121 VQONGVETSVLSDQEARLQSIDPEGRDKFVFTGGRGAGHAMVTVASDITEARQRIL 180
DB 995 IGNASGNAD-----AKVTFD-KVKDSKISTDGHN-----VTLNSEV----- 1031
QY 181 ELLEPKGTGSKGAGESKGVGELRESNSGAENTTETQTSSTSSLRSDPKLWALGTAVT 240
DB 1032 -----KTSNGSSNAGNDNSTG-----LTISAKQVTVNNVNTSHKTI-----NISAAAGNVTT 1078
QY 241 --GLIGLAATGIVQALALTPEDSPPTTDPDAAASATETATRDQLTK- AFON----- 290
DB 1079 KEGTINATTGSEVTA-----QNGTIKGNITSQNVTVTATENLVTTENAVINATSGTVN 1133
QY 291 -----PDNQKNIDELGNAIP-SGVLKDDVVANIEEQAKAAGEAKQOAIEN 336
DB 1134 ISTKTDGIRKGIESTSGNVNITASGNTLKVSNITGQDVTVTADAGALTTTAGSTISATTG 1193
QY 337 NAAQAQKYYDEQAQRQEELKYSVGA-GYGLSGALILGGGIG---VAVTAALHRKNQVQEQ 392
DB 1194 NANITTKTGIDNGK-----VESSSGSVTLVATGATLAVGNISGNVTVITADSGKLTSTVGS 1249
QY 393 TTTTITTTTTSAR-----TVENKPNANTPAQGNVDTPGSEDTMESRR----- 435
DB 1250 TINGINSVTTSQSGDIECTISGNTVNVTAFTGDL-TIGNSAKVEAKNGAATLTAESGKL 1308
QY 436 -----SSMASTSTFTDSSIGGPCRIRMLKHKRMRRCRLLILIRLFRWIGIQ 486
DB 1309 TQTGSSITSSNGQTTLTAKD-SSIAG-----N 1335
QY 487 ISWVYSTIQHPPTDNGARLLGNPSAGIQSTYARLALSGGLRHDGGLTGCSNAVNT 546

Db 1336 INAAVTLNTGTLTTGDSKI--NATSGTLTINAKDA-----KLDGAASGDRVTVNA 1386
QY 547 SNNPPAPGS 555
Db 1387 TN---ASGS 1392
RESULT 5
US-08-719-641-10
; Sequence 10, Application US/08719641
; Patent No. 6218141
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen J
; TITLE OF INVENTION: High Molecular Weight Surface Proteins
; TITLE OF INVENTION: of No. 6218141-Typeable Haemophilus
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Mattare, Ltd.
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; STREET: Bldg. 1
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/719,641
; FILING DATE:
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/302,832
; FILING DATE: 16-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US PCT/US93/02166
; FILING DATE: 16-MAR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9205704.1
; FILING DATE: 16-MAR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Berkstresser, Jerry W
; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: 1038-625
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0813
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1529 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-719-641-10

Query Match 4.4%; Score 124.5; DB 4; Length 1529;
Best Local Similarity 20.2%; Pred. No. 0.041;
Matches 123; Conservative 76; Mismatches 233; Indels 177; Gaps 28;
QY 26 TDGAGG-----RGQLNSTGPGLSRALETPVRNSMADS-----GDNRASDVFCPLPVPNMR 75
Db 882 TTNASGTQKTIINGNTNEKGLN-----IKNIKADAEIOIGGNISQKEGNLTSSDK 934
QY 76 L-AASETLNDGFE-VLHDGHPDITLNROIGSSVFRVE-----TQEDCKHTA 120
Db 935 VNITNQITIKAGVEGRSSSEAEANLTIQTKEKLAGDLNLSGPNKAEITAKNGSDLT 994
QY 121 VGORNGVETSVLSDQEARLQSIDPEGKDFVFTGGRGAGHAMVTVASDITEARQRL 180

Db 995 IGNASGNAD-----AKVTFD-KVKDSKISTDGHN-----VTLNSEV----- 1031
QY 181 ELLEPKGTGESKGAGSKGVGELRESNSGAENTETQTSTSTSSLRSDPKMLALGTVAT 240
Db 1032 -----KTSNGSSNAGNDNSTG---LTSIAKDVTVNNNVTSHKTI-----NISAAAAGNVTT 1078
QY 241 --GLIGLAATGIVOALALTPEDSPPTTDPDAAASATETATRDOLTKF-AFON----- 290
Db 1079 KEGTTINATTGSVEVTA-----QNGTIKGNITSQNVTVTATENLVTTENAVINATSGTVN 1133
QY 291 -----PDNOKYNIDELGNAIP-SGVLKDDVVANIEEQAKAAGEEAKQQAIEIN 336
Db 1134 ISTKTGDIKGGIESTSGNVNITASGNTLVKSVNITGQDVTVTADAGALTTAGSTISATTG 1193
QY 337 NAAQAKYDEQQAQROELKVVSSGA-GYGLSGALIIGGGIG---VAVTAALHRKNOPVEQ 392
Db 1194 NANITTKTGIDNGK-----VSSSGSVTLVATGATLAVGNISGNTVTITADSGKLTSTVGS 1249
QY 393 TTTTTTTTTTTTSAR-----TVENKPNANTPAQGNVDTPGSEDTMESRR----- 435
Db 1250 TINGTNSVTSSQSGDIEGTISGNTVNVNTASTGDL-TIGNSAKVEAKNGAATLTASGKL 1308
QY 436 -----SSMASTSTFPDTSIGGPCRIIRMLKLRHRCMIRRCRLILIRLRIWGIQ 486
Db 1309 TTQGTSSITSSNGQVTLTAKD-SSIAG-----N 1335
QY 487 ISVYVSTIQHPDRTDNGARLLGNPSAGIOSTYARLALSGGLRHDGGLTGCSNSAVNT 546
Db 1336 INAAVTLNTGTLTTGDSKI--NATSGTLTINAKDA-----KLDGAASGDRVTVNA 1386
QY 547 SNNPPAPGS 555
Db 1387 TN---ASGS 1392
RESULT 6
US-08-617-697-10
; Sequence 10, Application US/08617697
; Patent No. 5977336
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen J
; TITLE OF INVENTION: High Molecular Weight Surface Proteins
; TITLE OF INVENTION: of No. 5977336-Typeable Haemophilus
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Mattare, Ltd.
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; STREET: Bldg. 1
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/617,697
; FILING DATE: 01-APR-1996
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/302,832
; FILING DATE: 05-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US PCT/US93/02166
; FILING DATE: 16-MAR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Berkstresser, Jerry W
; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: 1038-557
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810

TELEFAX: (703) 415-0813
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1600 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-617-697-10

Query Match 4.4%; Score 124.5; DB 2; Length 1600;
Best Local Similarity 20.2%; Pred. No. 0.044;
Matches 123; Conservative 76; Mismatches 233; Indels 177; Gaps 28;

QY 26 TDGAGG-----RGQLNSTGPLSRALFTVPVNSMADS-----GDNRASDPGLPVNPMR 75
DB 952 TTNASGTQKTINGNITKEGDLN-----IKNIKADAEIQIGNISQKREGNLTISDDK 1004
QY 76 L-AASEITLNDGFE-VLHDHGPLOTNLROIGSSVFRVE-----TOEDGKHIA 120
DB 1005 VNINQIIRKAGVEGRSDSEAEANANLTQTKELKAGDLNISGFNKAETAKNGSLT 1064
QY 121 VQORNGVTSVVLSDQYARLQSIDPEGKDFVFTGGRGGAGHAMVTVASDITEARQRIL 180
DB 1065 IGNASGGNAD-----AKKVTED-KVKDSKISTDGHN-----VTLNSEV----- 1101
QY 181 ELLEPKGTGKSGKAGESKVGELRESNGAENTTQTSTSSLSRDPKWLALGTAT 240
DB 1102 -----KTSNGSSNAGNDNSTG-----LTISAKDVTVNNVNTSHKTI-----NISAAGNVT 1148
QY 241 -GLIGLAATGVQALALTPEPDSPITTTDPDAAASATETATRDQLTKE-AFQN----- 290
DB 1149 KEGTINATGSEVTA-----QNGTIKGNITSQNVTVTATENLVTTENAVINATSGTVN 1203
QY 291 -----PDNQKVNIDELGNAIP-SGVLKDDVYVANIEQAAKAAEAKQQAIEA 336
DB 1204 ISTKTDIKGGIESISGNVITASGNLTKVSNIQTQDVTVTADAGALTTAGSTISATTG 1263
QY 337 NAAQAKKYDEQAKRQBELKVSSGA-GYGLSGALILGGIG-----YAVTAAHLRKNOPVEQ 392
DB 1264 NANITTKTGDINGK-----VESSSGSVTLVATCATLAVGNISGNTVTITADSGKLTSTVGS 1319
QY 393 TTTTITTTTTTSAR-----TVENKPNANTPAQGNVDTPGSEDTMESRR----- 435
DB 1320 TINGTNSVTTSSQSDIGETISGNTVNTVNTAGTDL-TIGNAKVEAKNGAATLTAESGL 1378
QY 436 -----SSMASTSTFTFDTSITSGGPCIRIMLMKHKRMIRCRLLILRLFRINGIQ 486
DB 1379 TTQTGSSITSSNGQTLTAKD-SSIAG-----N 1405
QY 487 ISVYVSTIQHPRTDTDNGARLLGNPSAGIOSTYARLALSGLRHDMGGLTGGSNSAVNT 546
DB 1406 INAAVNTLTGTTLTTGDSKI--NATSGTLTINAKDA-----KLDGAASGDRTVVNA 1456
QY 547 SNNPAPGS 555
DB 1457 TN---ASGS 1462

RESULT 7
US-08-409-995-5
; Sequence 5, Application US/08409995
; Patent No. 5646259
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen I.
; APPLICANT: St. Geme III, Joseph W.
; TITLE OF INVENTION: Haemophilus Adhesion Proteins
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
; STREET: Four Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: CA

COUNTRY: USA
ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/409,995
FILING DATE: 24-MAR-1995
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Silva, Robin M.
REGISTRATION NUMBER: 38,304
REFERENCE/DOCKET NUMBER: A-61053/RFT
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
TELEX: 910 277299
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 658 amino acids
TYPE: amino acid
STRANDEDNESS: double
TOPOLOGY: unknown
US-08-409-995-5

Query Match 4.3%; Score 123; DB 1; Length 658;
Best Local Similarity 22.4%; Pred. No. 0.015;
Matches 111; Conservative 58; Mismatches 171; Indels 156; Gaps 25;

QY 10 PNVNNSIPPAPLPESQTDG-----ACGRQLNSTGPLSRALFTVPVNR--SMADSGD 60
DB 194 PKVN-----VTSTDGLKFAQDAAGANGDTTVHLNGIGSTLTDTLVGSPATHIDGGD 245
QY 61 -----NRASDV-----PGLPVNPMRLAASEITLNDGFEVLHDHGPDLTLNRQIGSSVFRV 110
DB 246 QSTHYTRAASIKDVLNAGWNKGVK-AGSTTGQSENDFVHTYDVEFLSADTETTTTV 304
QY 111 ETQEDGKHIAVGQRNGVTSVVLSDQYARLQSIDPEGKDFVFTGGRGGAGHAMVTVAS 170
DB 305 DSKENGKRTEV--KIGAKTSVI-----KEKDGKLTG----- 334
QY 171 DITEARQRILELLEPKGTGKSGKAGESKVGELRESNGAENTTQTSTSSLSRDPK 230
DB 335 ---KANK---ETNKVDGANATEDADEGKLVAKDVIDAVNKTGWRIKTTDANGONGD-- 386
QY 231 LWLALGTVATGLIGLAATGIVQALALTPEPDSPIT-----TOPDAAASATETAT 279
DB 387 -----PATVASGNTVTFASGNGTTATVNGTGTITVYKDYAKVGDGLKLDGDKTAADT-TA- 440
QY 280 RDQLTKFAQNPNDOKVNIDELGNALPISGVLKDDVYVANIEEQ-----AK----- 323
DB 441 ---LTVNDGNKANN-----PRGKVAD--VASTDEKKLVTAAGLVLTALNSLSWT 483
QY 324 AAGEBAKQQAIENTNNAQAQ--KKYDEQAKRQBELKV-SSGAG--YGLSGALILGGIGVA 378
DB 484 TTAEEADGGTLDGNASEQEVKAGDKVTFKAGKNLKVKEGANFTYSLQDALT---GL--- 537
QY 379 VTAALHRKNQPVQETTTTTTTTTTTSARTVENKPA--NNTPAQGNVDTPGSEDTMESRRSS 437
DB 538 -----TSITLGTGNNGAKTEINKDGLTITPANG-----AGAN 569
QY 438 MASTSSTFTDTSISGG 453
DB 570 NANTISVTKDGISAGG 585

RESULT 8
US-08-685-467-5
; Sequence 5, Application US/08685467
; Patent No. 6060059

GENERAL INFORMATION:
APPLICANT: St. Gene III, Joseph W.
APPLICANT: Barenkamp, Stephen J.
TITLE OF INVENTION: HAEMOPHILUS ADHESION PROTEINS
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Flehr, Hobbach, Test, Albritton & Herbert
STREET: Four Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: United States
ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/685,467
FILING DATE: 22-JUL-1996
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/409,995
FILING DATE: 24-MAR-1995
ATTORNEY/AGENT INFORMATION:
NAME: Silva, Robin M.
REGISTRATION NUMBER: 38,304
REFERENCE/DOCKET NUMBER: A-61053-2/RFT/RMS
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
TELEX: 910 277299
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 658 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-685-467-5

Query Match 4.3%; Score 123; DB 3; Length 658;
Best Local Similarity 22.4%; Pred. No. 0.015;
Matches 111; Conservative 58; Mismatches 171; Indels 156; Gaps 25;
QY 10 PWNNSIPPAPPLPSQTDG-----AGRGQLNSTGPGLSRALFTVPVRN--SMADSGD 60
DB 194 PKVN-----VTSTDTGLKFAKDAAGANGDTTVHLNGIGSTLTDTLVGSPATHIDGGD 245
QY 61 -----NRASDV-----PGLPVNPMRLAASEITLNDGFVLDHGHPLDTLNRQIGSSVFRV 110
DB 246 QSTHTYRAASIKDVLNAGWNKGVK-AGSTTGQSENVDVHTYDVEFLSADTETTTTV 304
QY 111 ETQEDGKHIAVGQNGVTSVVLSDQEVARLQSIDPEGKDFVFTGGRGGAGHAMVTVAS 170
DB 305 DSKENGKRTEV--KIGAKTSVI-----KERDGLFTG----- 334
QY 171 DTEARQRIELLEPGKTESKAGESKVGELRESNSGAENTTETQTSTSTSLRSDPK 230
DB 335 ---KANK-----ETNKVDGANATEDGKGLVTAQVIDAVNKTGWRIKTYTDANGONGD-- 386
QY 231 LWLALGTVAATGILGALATGIVQALATPPDPSFTT-----TDPDAASATETAT 279
DB 387 ----FATVAGTNVTFASNGTATVNTNGDTGITYKYDAKVGDLKLDGDKLAADT-TA- 440
QY 280 RQLTKEAFONPDNOKNIDELGNALPSVLKDDVYANTEEQ-----AK----- 323
DB 441 ---LTVDNCKNANN-----PKGVAD--VASTDEKKLVAKGLVTAALNSLWT 483
QY 324 AGEERAKQQAIEENNAQ--KKYDEQAKRQELKV--SSGAG--YGLSALLILGGGIGVA 378
DB 484 TTAEEADGGTLDGNASEVEQKAGDKVTFRAGNKLKVKQEGANFTYSLQDALI---GL--- 537

QY 379 VTAALHRKNQPVQEOTTTTTTTTTTTTTTTSARTVENKPA-NNTPAGNVDTFGSEDTWESRSS 437
DB 538 -----TSITLGTGNGNGAKTEINKDGLTITPANG-----AGAN 569
QY 438 MASTSTSTFFDSSIGG 453
DB 570 NANTISVTKDGISAGG 585
RESULT 9
US-08-913-942-5
; Sequence 5, Application US/08913942
; Patent No. 6200578
; GENERAL INFORMATION:
; APPLICANT: St. Gene, Joseph
; APPLICANT: Barenkamp, Stephen J.
; TITLE OF INVENTION: HAEMOPHILUS ADHESION PROTEINS
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Flehr Hobbach Test Albritton & Herbert LLP
; STREET: Four Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: United States
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/913,942
; FILING DATE: 29-DEC-1997
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/409,995
; FILING DATE: 24-MAR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US96/4031
; FILING DATE: 22-MAR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Vance, Dolly A.
; REGISTRATION NUMBER: 39,054
; REFERENCE/DOCKET NUMBER: A-61053-1/RFT/RMS/DAV
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 658 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; US-08-913-942-5

Query Match 4.3%; Score 123; DB 4; Length 658;
Best Local Similarity 22.4%; Pred. No. 0.015;
Matches 111; Conservative 58; Mismatches 171; Indels 156; Gaps 25;
QY 10 PWNNSIPPAPPLPSQTDG-----AGRGQLNSTGPGLSRALFTVPVRN--SMADSGD 60
DB 194 PKVN-----VTSTDTGLKFAKDAAGANGDTTVHLNGIGSTLTDTLVGSPATHIDGGD 245
QY 61 -----NRASDV-----PGLPVNPMRLAASEITLNDGFVLDHGHPLDTLNRQIGSSVFRV 110
DB 246 QSTHTYRAASIKDVLNAGWNKGVK-AGSTTGQSENVDVHTYDVEFLSADTETTTTV 304
QY 111 ETQEDGKHIAVGQNGVTSVVLSDQEVARLQSIDPEGKDFVFTGGRGGAGHAMVTVAS 170

Db 305 DSKENGKRETEV--KIGAKTSVI-----KEKDGKLTG----- 334
Qy 171 DITEARQILELLEPKGTGSGKAGESKGVGELRESNSGAENTTETQTSTSSLSRDPK 230
Db 335 ---RANK---ETNKVDGANATEDAGKGLVAKDVDAVNKTGWRIKTTDANGONGD-- 386
Qy 231 LWALGTVATGLIGLAATGIVQALALATPEPDSPTT-----TPDAAAASATETAT 279
Db 387 ---FATVASGTNVTFFASGNGTTATVTNGTDGIVKYDAKVGDLGLKDGKIAADT-TA- 440
Qy 280 RDQLTKEAFQPNQKVNIDELGNAIPSGVLKDDVVAIEQO---AK----- 323
Db 441 ---LTVNDGKNANN-----PKGVAD--VASTDEKKLVTAKGLVTALNSLSWT 483
Qy 334 AAGEAKQQAIAENNAQAQ--KKYDEQQAQROBELKV--SSGAG--YGLSGALILGGIGVA 378
Db 484 TTAADGGLTLDGNASEQEVKAGDKVTFKAGKNLKVQEGANFTYSLQDALT---GL--- 537
Qy 379 VTAALHRKNQVEQTTTTTTTTTTTTTSARTVENKPA--NNTPAQGNVDTPGSEDTMESRRSS 437
Db 538 -----TSITLGTGNGNKAETKNGDGLTITPANG-----AGAN 569
Qy 438 MASTSSTFFDTSSIGG 453
Db 570 NANTISVTKDGISAGG 585

RESULT 10

US-08-409-995-2
; Sequence 2, Application US/08409995
; Patent No. 5646259
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen I.
; APPLICANT: St. Geme III, Joseph W.
; TITLE OF INVENTION: Hemophilus Adhesion Proteins
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
; STREET: Four Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/409,995
; FILING DATE: 24-MAR-1995
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Silva, Robin M.
; REGISTRATION NUMBER: 38,304
; REFERENCE/DOCKET NUMBER: A-61053/RFT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1098 amino acids
; TYPE: amino acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
US-08-409-995-2

Query Match 4.38; Score 123; DB 1; Length 1098;
Best Local Similarity 22.4%; Pred. No. 0.033;
Matches 111; Conservative 58; Mismatches 171; Indels 156; Gaps 25;

Qy 10 PNVNNSIPPAPPLPSQTDG-----AGRGQLNSTGPLGSRALFTVRN--SMADSGD 60
Db 194 PKVN-----VTSTDLKFAKDAAGANGDTTVHLNGICSTLTDTLVGSPTHIDGSD 245
Qy 61 -----NRASDV-----PGLPVNPMRLAASEITLNDGFVHLHDHGLDPLNRLQIGSSVFRV 110
Db 246 QSTHYTRAASIKDVLNAGWNKGVK-AGSTTQCSNVDFVHTYDTVEFLSADTETTTTV 304
Qy 111 ETQEDGKHIAVGQNGVETSVLSDOEYARLQSIDPECKDKFVFTGGRGGAGHAMVTVAS 170
Db 305 DSKENGKRETEV--KIGAKTSVI-----KEKDGKLTG----- 334
Qy 171 DITEARQILELLEPKGTGSGKAGESKGVGELRESNSGAENTTETQTSTSSLSRDPK 230
Db 335 ---RANK---ETNKVDGANATEDAGKGLVAKDVDAVNKTGWRIKTTDANGONGD-- 386
Qy 231 LWALGTVATGLIGLAATGIVQALALATPEPDSPTT-----TPDAAAASATETAT 279
Db 387 ---FATVASGTNVTFFASGNGTTATVTNGTDGIVKYDAKVGDLGLKDGKIAADT-TA- 440
Qy 280 RDQLTKEAFQPNQKVNIDELGNAIPSGVLKDDVVAIEQO---AK----- 323
Db 441 ---LTVNDGKNANN-----PKGVAD--VASTDEKKLVTAKGLVTALNSLSWT 483
Qy 334 AAGEAKQQAIAENNAQAQ--KKYDEQQAQROBELKV--SSGAG--YGLSGALILGGIGVA 378
Db 484 TTAADGGLTLDGNASEQEVKAGDKVTFKAGKNLKVQEGANFTYSLQDALT---GL--- 537
Qy 379 VTAALHRKNQVEQTTTTTTTTTTTTTSARTVENKPA--NNTPAQGNVDTPGSEDTMESRRSS 437
Db 538 -----TSITLGTGNGNKAETKNGDGLTITPANG-----AGAN 569
Qy 438 MASTSSTFFDTSSIGG 453
Db 570 NANTISVTKDGISAGG 585

RESULT 11
US-08-685-467-2
; Sequence 2, Application US/08685467
; Patent No. 6060059
; GENERAL INFORMATION:
; APPLICANT: St. Geme III, Joseph W.
; APPLICANT: Barenkamp, Stephen J.
; TITLE OF INVENTION: HAEMOPHILUS ADHESION PROTEINS
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
; STREET: Four Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: United States
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/685,467
; FILING DATE: 22-JUL-1996
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/409,995
; FILING DATE: 24-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Silva, Robin M.
; REGISTRATION NUMBER: 38,304
; REFERENCE/DOCKET NUMBER: A-61053-2/RFT/RMS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEX: (415) 398-3249
; TELEX: 910 277299

INFORMATION FOR SEQ ID NO: 2:	
SEQUENCE CHARACTERISTICS:	
LENGTH: 1098 amino acids	
TYPE: amino acid	
STRANDEDNESS: unknown	
TOPOLOGY: unknown	
MOLECULE TYPE: DNA (genomic)	
US-08-685-467-2	
Query Match	
Best Local Similarity 4.3%; Score 123; DB 3; Length 1098;	
Matches 111; Conservative 58; Mismatches 171; Indels 156; Gaps 25;	
QY	10 PNVNNSIPPAPPLPSOTDG-----AGGRGQLINSTGPLGSRALFTPVNR--SMADSGD 60
DB	194 PKVN-----VTSTTDLGKFAKDAAGANGDTTVHLNGIGSTLTDLTVGSPATHIDGGD 245
QY	61 -----NRASDV-----PGLPVPNPMRLAASEITLNDGFEVLHDHGPLDNLNRQIGSSVFRV 110
DB	246 QSTHYTRAASIKDVLNAGWNKGVK-AGSTTGQSENVDVFVHTYDTVEFLSADTETTIVTV 304
QY	111 ETQEDGKHIAVGQNGVETSVLSQDEYARLQSIDPEGDKFVFTGGRGGAGHAMVTVAS 170
DB	305 DSKENGKRTVEV--KIGAKTSVI-----KEKDGKLFVG----- 334
QY	171 DITAEARQRIELLEPEKGTGESKGAGESKGVGELRESNGAENTTETOTSTSTSLSDPK 230
DB	335 ---KANK---ETNKVDGANATEDADEGKGLVAKVDIDAVNKTGWRIKTTDANGQNGD-- 386
QY	231 LWALGTVATGLIGLAATGIVQALALTPEDPSPTT-----TDPDAASATETAT 279
DB	387 ----FATVASTGNTVFASGNGTTATVTNGTIGITVKYDAKVGDGLKLDGDKIAADT-TA- 440
QY	280 RDOLTKAEFQNDQKVNIDELGNAIPSGVLKDDVVAIEEQ-----AK----- 323
DB	441 ---LTVNDGKNANN-----PKGKVAD--VASTDEKLVAKGLVTALNSLSWT 483
QY	324 AAGEBAKQQAIAENNAQAQ--KKYDEQQAQKROELKV--SSGAG--YGLSGALILGGGIGVA 378
DB	484 TTAEEADGGTLDGNASEQEVKAGDKVTFKAGKNLKVQEGANFTYSLODALT---GL--- 537
QY	379 VTAALHRKNQPVQETTTTTTTTTTTTTSARTVENKPA--NNTPAQGNVDTPGSEDTMESRRSS 437
DB	538 -----TSITLGTGNGNGAKTEINKDGLTITPANG-----AGAN 569
QY	438 MASTSTSTFFDTSSIGG 453
DB	570 NANTISVTKDGISAGG 585
RESULT 13	
US-08-913-942-2	
Sequence 2, Application US/08913942	
Patent No. 6200578	
GENERAL INFORMATION:	
APPLICANT: St. Geme, Joseph	
APPLICANT: Barenkamp, Stephen J.	
TITLE OF INVENTION: HAEMOPHILUS ADHESION PROTEINS	
NUMBER OF SEQUENCES: 19	
CORRESPONDENCE ADDRESS:	
ADDRESSEE: Flehr Hobbach Test Albritton & Herbert LLP	
STREET: Four Embarcadero Center, Suite 3400	
CITY: San Francisco	
STATE: California	
COUNTRY: United States	
ZIP: 94111-4187	
COMPUTER READABLE FORM:	
MEDIUM TYPE: Floppy disk	
COMPUTER: IBM PC compatible	
OPERATING SYSTEM: PC-DOS/MS-DOS	
SOFTWARE: PatentIn Release #1.0, Version #1.30	
CURRENT APPLICATION DATA:	
APPLICATION NUMBER: US/08/913,942	
FILING DATE: 29-DEC-1997	
CLASSIFICATION: 514	
PRIOR APPLICATION DATA:	
APPLICATION NUMBER: US 08/409,995	
FILING DATE: 24-MAR-1995	
PRIOR APPLICATION DATA:	

APPLICATION NUMBER: PCT/US96/4031
FILING DATE: 22-MAR-1996
ATTORNEY/AGENT INFORMATION:
NAME: Vance, Dolly A.
REGISTRATION NUMBER: 39,054
REFERENCE/DOCKET NUMBER: A-61053-1/RFT/RMS/DAV
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
TELEX: 910 277299
CURRENT APPLICATION DATA:
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1098 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-913-942-2

Query Match 4.3%; Score 123; DB 4; Length 1098;
Best Local Similarity 22.4%; Pred. No. 0.033;
Matches 111; Conservative 58; Mismatches 171; Indels 156; Gaps 25;

QY 10 PNVNNSIPPAPPLPSQTDG-----AGRGOLINSTGLSGSRALFTVPVN--SMADSGD 60
DB 194 PKVN-----VSTTDGLKFAKDAAGANGDTTVHLNGIGSTLTLTLVGSPTHIDGSD 245
QY 61 -----NRASDV-----PGLPNPMLAASEITLNDGFVHLHDGPDLTNRQIGSSVFRV 110
DB 246 QSTHYTRAASIKDVLNAGWNKGVK-AGSTTGQSENVDFVHTYDVFELSADTETTTTV 304
QY 111 ETQDGKHIAGVQRNGVTSVLSDOEYARLOSIDPEGKDFVFTGGRGAGHAMVTVAS 170
DB 305 DSKNGKRTVE--KIGAKTSVI-----KEKDGKLEFG----- 334
QY 171 DITEARQILELLEPKGTGSGKAGESKGVGELRESNGAENTTQTSTSLRSDPK 230
DB 335 ---KANK---ETNKVDGANATEDAGEGLTAKDVIDAVNKTGHRKITTANGONGD-- 386
QY 231 LWALGTVATGLIGLAATGIVQALALTPPEPSPTT-----TPDAAAASATETAT 279
DB 387 ---PATVASGTNTVFASGNGTTATVTNGTDGTTVKYDAKVGDKLGDGKIAADT-TA- 440
QY 280 RDQLTKEAFQNDPNQKVNIDELGNALPSGLKDDVVANIEEQ-----AK----- 323
DB 441 ---LTVNDGKNANN-----PKGVAD---VASTDEKLVAKGLVYALNLSWT 483
QY 324 AAGERAKQQAIEENNAQA--KKYDQQAQKROEELKV-SSGAG--YGLSGALITLGGGIGA 378
DB 484 TTAADAGTLDGNASEQEVKAGDKVTEKAGNKLKVKQEGANFYSLQDALT---GL--- 537
QY 379 VTAALHRKNQVEQGTTHTTTTTTSARTVKNKPA-NNTPAQGNVDTPGSEDTMESRRSS 437
DB 538 -----TSTLTGNGNGAKTEINKDGLTITTPANG-----AGAN 569
QY 438 MASTSSSTFTDSSIGG 453
DB 570 NANTISVTKDGISAGG 585

RESULT 14
US-08-296-791-6
Sequence 6, Application US/08296791
Patent No. 6245337
GENERAL INFORMATION:
APPLICANT: St. Gene III, Joseph W.
APPLICANT: Falkow, Stanley
TITLE OF INVENTION: Haemophilus Adherence and Penetration
TITLE OF INVENTION: Protein
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Flehr, Hobbach, Test, Albritton & Herbert

STREET: 4 Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: United States
ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/296.791
FILING DATE: 25-AUG-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Trecartin, Richard F.
REGISTRATION NUMBER: 31,801
REFERENCE/DOCKET NUMBER: A-59941/RFT/RMS
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
TELEX: 910 277299
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 1848 amino acids
TYPE: amino acid
TOPOLOGY: unknown
US-08-296-791-6

Query Match 4.3%; Score 123; DB 4; Length 1848;
Best Local Similarity 21.3%; Pred. No. 0.076;
Matches 103; Conservative 52; Mismatches 174; Indels 154; Gaps 23;
QY 17 PPAP-----PLPSQTDGAGRGGLINSTGLSGSRALFTVPVNSMA 56
DB 1038 PPAPATAISAEQETPRPAETAQAPEMETNTANS-----TETAPKSDTATQTEPNSES 1092
QY 57 DSGDNRAADVGLPNPMLA-----ASEITLNDGFVHLHDGPDLTNRQIGSSVFRVE 111
DB 1093 VPSETTEKVAENPPQENETVAKNEQEAETPTPQNG-EVAKEDQPTVEANTQTNEA----- 1146
QY 112 TQEDGKHIAGVQRNGVTSVLSDOEYARLOSIDPEGKDFVFTGGRGAGHAMVTVASD 171
DB 1147 TQSEK-----TEETQTAETKSEPTES-----VTVSEN 1174
QY 172 ITEARQILELLEPKGTGSGKAGESKGVGELRESNGAENTTQTSTSSLSRSDPKL 231
DB 1175 -----QPEKTVSQ--STEDKVVVE-KEEKAKVE-TEETQKAPQVTS-KEPPK- 1216
QY 232 WLALGTVATGLIGLAATGIVQALALTPPEDS-PTTDDPDA-----ASATETATR 280
DB 1217 -----QA-----EPAPEEVPDTNAAEEAQAALQOQPTTVAEAETTS 1253
QY 281 DQLTKEAFQNDPNQKVNIDELGNALPSGLKDDVVANIEEQAKAAGEAKQQAIEENNAQA 340
DB 1254 NSKPAEETQOP-SEKTNAPVTPPVSENTATQP--TEETEAKVE-KEKQVPPQVSAQE 1309
QY 341 OKKYDEQQAQKROEELKVSSGAGYLSGALILGGGIGVAVTAALHRKN-----QPVEQTITT 396
DB 1310 SPKQEQPAKPAQATKPAEP-----ARENVLTKNVGEPQPAQPOQ 1353
QY 397 TTTTTSARTVENKPA-----NNTPAQGNVDTPGSEDTMESRRSSMASTSTFTD 447
DB 1354 STAVPTTGTAAANSKPAKPAQAKPQTEPARENVSTVNTKEP-QSOTSATVSTEQPAKE 1412
QY 448 TSS 450
DB 1413 TSS 1415
RESULT 15
PCT-US95-10661A-6

Thu Sep 27 15:29:55 2001

Sequence 6, Application PC/TUS9510661A
GENERAL INFORMATION:
APPLICANT: Washington University, et al.
TITLE OF INVENTION: Haemophilus Adherence and Penetration Protein
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Flehr, Hobbach, Test, Albritton & Herbert
STREET: 4 Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: United States
ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/10661A
FILING DATE: 16-AUG-1995
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/296,791
FILING DATE: 25-AUG-1994
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Trecartin, Richard F.
REGISTRATION NUMBER: 31,801
REFERENCE/DOCKET NUMBER: FP-59941/RFT
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
TELEX: 910 277299
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 1848 amino acids
TYPE: amino acid
TOPOLOGY: unknown
PCT-US95-10661A-6

Query Match 4.3%; Score 123; DB 5; Length 1848;
Best Local Similarity 21.3%; Pred. No. 0.076;
Matches 103; Conservative 52; Mismatches 174; Indels 154; Gaps 23;

QY	17	PPAP-----PLPSQTDGAGRGGLINSTGPLGSRALFTPVNRNSMA 56
DB	1038	PPAPATESAIASEQPETRAPETAQPAEETNTANS-----TETAPKSDTATQTENPNSES 1092,
QY	57	DSGDNRASDPGLVPNPMRLA-----ASEITLNDGFEVLHDHGPPLDTLNRIGSSVFRVE 111
DB	1093	VPSEETKVAENPPQENETVAKNEQEATEPTPNG-EVAKEDQPTVEANTQTNEA-----1146
QY	112	TQEDKGHTAVGQNGVTSVVLSDQEIARLASIDPEGKDKFVFTGGRGAGHAMVTVASD 171
DB	1147	TQSEKG-----TEETQTATKSEPTES-----VTVSEN 1174
QY	172	ITEARQRIULELLEPKGSGKAGSGVGELESNGAENTTETOTSTSSLSRSDPKL 231
DB	1175	-----QPKTVSQ--STEDKVVEE-KEEKAKVE-TEETQKAPQVTS-KEPK- 1216
QY	232	WLALGTAVATGLIGLAATGIVQALALTPEDS-PTTTPDDAA-----ASATETATR 280
DB	1217	-----QA---EPAEEVPTDTNAAEQALQOQTPTTVAEAETTSP 1253
QY	281	DOLTKEAFONPNQKNIDELGNALPSGVLKDDVVANTEEQAQAKGEEAKQAQAIENNAQA 340
DB	1254	NKPAEETQOP-SEKTNAEPTVPVSENATQP--TETEETAKVE-KETQEVPPVASOE 1309
QY	341	QKYDQQAQKROBELKLVSSGAGYGLSGALILGGIGGVAVTAALHRKN-----QPVEQTTT 396
DB	1310	SPKQEQPAKPAQATKPAEP-----ARENVLTTKNVGEPPQAPQPTQ 1353

QY	397	TTTTTTTSARTVENKPA-----NNTPAQGNVDTPGSEDYTMESRRSSMASTSTTFPD 447
DB	1354	STAVPTTGETAANSKPAKPAQAKPQTEPARENVTNTKEP-QSQTSATVSTEQPAKE 1412
QY	448	TSS 450
DB	1413	TSS 1415

Search completed: September 27, 2001, 14:21:08
Job time: 44 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 27, 2001, 14:21:42 ; Search time 25.62 Seconds
(without alignments)
1662.044 Million cell updates/sec

Title: US-09-189-415a-4

Perfect score: 2851

Sequence: 1 MPIGNLGHNPVNNNSIPPAP.....SNSAVNTSNPPAGSHRFV 559

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR_68:*

1: pir1.*

2: pir2.*

3: pir3.*

4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2624.5	92.1	558	2 E86045	probable transloca
2	170	6.0	1229	2 T25697	hypothetical prote
3	165.5	5.8	2232	2 T34434	hypothetical prote
4	162.5	5.7	1291	2 T13389	hypothetical prote
5	152.5	5.3	973	2 C85693	probable membrane
6	148.5	5.2	770	2 T51024	related to C2H2 zi
7	148	5.2	1192	2 T18611	probable serine/th
8	146.5	5.1	1122	2 G64887	probable tail fibe
9	144.5	5.1	3507	2 T34513	hypothetical prote
10	143.5	5.0	1275	2 T33369	hypothetical prote
11	143.5	5.0	1630	2 A53577	ascites sialoglyco
12	142	5.0	563	2 A36054	mucin homolog - bo
13	141.5	5.0	1962	2 A32634	lactocepin (EC 3.4
14	140.5	4.9	1026	1 A40315	maternal effect pr
15	140.5	4.9	1829	2 T24583	hypothetical prote
16	139	4.9	2468	2 A83412	hypothetical prote
17	137	4.8	837	2 T02761	outer arm dynein i
18	137	4.8	4558	2 C82199	RTX toxin RTXA VC1
19	136	4.8	1005	2 H85611	probable adhesin z
20	136	4.8	1005	2 H85663	probable adhesin z
21	134	4.7	2514	2 F81045	hemagglutinin/hemo
22	133.5	4.7	796	2 T21460	hypothetical prote
23	133	4.7	439	2 E71497	hypothetical prote
24	133	4.7	1589	2 T42233	submaxillary mucin
25	132.5	4.6	461	2 JN0097	secreted 45K prote
26	132.5	4.6	755	2 H86561	Cr456 hypothetical
27	132.5	4.6	755	2 B72061	hypothetical prote
28	132.5	4.6	1258	2 JQ0188	ice nucleation pro
29	132.5	4.6	1459	2 T32271	hypothetical prote

ALIGNMENTS

RESULT 1

E86045

probable translocated intimin receptor protein tir [imported] - Escherichia coli (str C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 31-Mar-2001

C:Accession: E86045

K:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; May
iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apoda
Nature 409, 529-533, 2001

A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.

A:Reference number: AB5480; MUID:21074935; PMID:11206551

A:Accession: E86045

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-558 <STO>

A:Cross-references: GB:AB005174; NID:gl2518449; PIDN:AAG58825.1; GSPDB:GN00145; UWGP:

A:Experimental source: strain O157:H7, substrain EDL933

C:Genetics:

A:Gene: tir

Query Match 92.1%; Score 2624.5; DB 2; Length 558;
Best Local Similarity 94.0%; Pred. No. 3.6e-148;
Matches 528; Conservative 3; Mismatches 24; Indels 7; Gaps 3;

Qy	1	MPIGNLGHNPVNNNSIPPAPLP	SQTGAGRGQLINSTG	PLGSRALFTPV	RNSMADSGD	60
Db	1	MPIGNLGHNPVNNNSIPPAPLP	SQTGAGRGQLINSTG	PLGSRALFTPV	RNSMADSGD	60
Qy	61	NRASDVPGLPVNPMRLAASEITL	NDGFEVLHDHGPDLTLNR	QIGSSVFRVETQ	EDGKHIA	120
Db	61	NRASDVPGLPVNPMRLAASEITL	NDGFEVLHDHGPDLTLNR	QIGSSVFRVETQ	EDGKHIA	120
Qy	121	VQQRNGVETSVVLSQDEYARLQ	SIDPEGRDKFVFTGGRGGAGHAMVT	VASDITEAR	QRIL	180
Db	121	VQQRNGVETSVVLSQDEYARLQ	SIDPEGRDKFVFTGGRGGAGHAMVT	VASDITEAR	QRIL	180
Qy	181	ELLEPKGTGESKAGESKVG	ELRESNGAENTTQTST	SSLSRSDPKL	LWALGTVAT	240
Db	181	ELLEPKGTGESKAGESKVG	ELRESNGAENTTQTST	SSLSRSDPKL	LWALGTVAT	240
Qy	241	GLIGLAATGIVQALALTPE	PDPSPTTTPDAAASATETAT	RDLQTK	EAQFONPNQKNVIDE	300
Db	241	GLIGLAATGIVQALALTPE	PDPSPTTTPDAAASATETAT	RDLQTK	EAQFONPNQKNVIDE	300
Qy	301	LGNALPSGVLKDDVVANIEEQ	AQKAGEAKQQA	IENNAQAKKYDEQ	AQKKEELKVSSG	360
Db	301	LGNALPSGVLKDDVVANIEEQ	AQKAGEAKQQA	IENNAQAKKYDEQ	AQKKEELKVSSG	360
Qy	361	AGYGLSALILGGGIGVAV	TAALHRKNQPVQTTTTTTT	TTTTTTT	SARTVENKPNANTPAQ	420
Db	361	AGYGLSALILGGGIGVAV	TAALHRKNQPVQTTTTTTT	TTTTTTT	SARTVENKPNANTPAQ	420

A: Introns: 238/3; 1225/1
A: Note: EG:115C2.10

Query Match 5.7%; Score 162.5; DB 2; Length 1291;
Best Local Similarity 20.6%; Pred. No. 0.075;
Matches 99; Conservative 61; Mismatches 189; Indels 131; Gaps 18;

QY 5 NLGHNPNVNSIPPAPPLPSQTDG---AGGRCQLINSTGPLGSRALFTTPVRNSMADSGDN 61
DB 438 NSTSNSNTNDSTGPSETSSNGLVSGAG-----GATGAAMLPTP---SQOSTGK 488
QY 62 RAS-----DVPGLPVNPMRLA-----ASEITLNDGFEVLHDHGPLDLNLRQ 102
DB 489 EATAAVALLEKLLPNVVVSPLTWKELRQKWKYDAEMIMANAQQOHHH----- 539
QY 103 IGSVFRVETQEDGKHIAVGRNGVETSVLVSDQRYARLOSIDPEGKDFVFTGGRGGAG 162
DB 540 --QHFFHHHHHHHHHGHQASTGAETAATAVQMAAMQKPG-----VGGTGAAG 588
QY 163 HAMVIVASDITEARQRIELLEPKGTGSGKAGSGVGLRESNSGAENTTETQTSTST 222
DB 589 NAGATTVSSVA-----AGAGSEVNGRSTSLRKS MRVNS-----TSSSI 627
QY 223 SSLSRDPKLLWALGTIVATGLIGLAATGIQVLAALALTPEDSPPTTTPDAAASATETATRDQ 282
DB 628 STASADEVI-----APVVAASISLPSKAPVVLMPCKPKPAOWATAALHQ 670
QY 283 LKPEAPQNDPNQKVIDELGNAPSGVLKDDVYANTEEQAAGAEAKOO-----A 333
DB 671 SQORQLRRSEROKELTD-----GESSD---TSSEQKK---EQKQDHLQPKMF 716
QY 334 IENNAQAQKDYDEQAQKROELKVSAGYGLSGALILGGIGVAVTAALHRKNQPVET 393
DB 717 LAEPQPEKS-EKQOEQOKRYTRNSAGRVGL-----VARLATAHNNN-----TA 760
QY 394 TTTTITTTTTSARTV---ENKPNANTPAQGNVDTPGSEDITMESRRSSMASTSTPFDTS 450
DB 761 TTYNSSSSNKATTITNCNNHNSNRHNSLSRLSVKSRKPAPSEASIPSTSS 820

RESULT 5
C85693
probable membrane protein of prophage CP-933X Z1918 [imported] - Escherichia coli (strain
C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 31-Mar-2001
C:Accession: C85693
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
killer, L.; Grothbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamoudis, K.; Apodaca,
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: C85693
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-973 <STO>
A:Cross-references: GB:AE005174; NID:g12514847; PIDN:AAG56007.1; GSPDB:GN00145; UWGP:Z19
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: Z1918

Query Match 5.3%; Score 152.5; DB 2; Length 973;
Best Local Similarity 22.1%; Pred. No. 0.2; Mismatches 169; Indels 83; Gaps 14;
Matches 87; Conservative 55

QY 79 SEITLNDGFEVLH-----DHGPLDLNROIGTSVFRVETQEDGKHIAVGO----- 123
DB 61 SVILLVEGPPSHAGTIVYENSQP-GTLDNLFGLM-----TEDDVRPEALRFELWVEE 114
QY 124 --RNGVETSVILSDQBYARLOSIDPEKDKFVFTGGRGAGHAMVTVASDITEARQRI 181
DB 115 VARN---ASAVAQNTAAAKKSASDAS-----TSAREATHA--TDAADSARA----- 156

QY 182 LLEPKGTGSKGSGKGVGLRESNSGAENTTETQTSTSSLSRDPKLLWALGTIVATG 241
DB 157 -----ASTSAGQAASQAQSSAGTASTKATEAKSAASAAESSK-----SAAAT 201
QY 242 LIILAATGIVQALALTPEDSPPTTTPDAAASATATATARDQLTKEAFQNPQKVNIDEL 301
DB 202 SAGAAKTSETNAVSOQSAATSTASTATTKASEAASSARDASASAKSAKSETSAAS---- 257
QY 302 GNAIPSGVLKDDVYANIEBQAAGAEAK-QQAIENNAQAQKDYDEQAQKROELKVS 360
DB 258 -----SASSAASSATAAGNSAKAAKTSETNAKSSSETAQAQSAAGSK--TA 303
QY 361 AGYGLSGALLGGIGVAVTAALHRKNQPVETTTTTTTTTTTTTSARTVENKPNANTPAQ- 419
DB 304 AALSASAASTSAQASASATAA-----GKSAESAASASTATTKAGENTEQAASAAASASA 359
QY 420 ---GNVDTPGSEDITMESRRSSMASTSTPFDTS 450
DB 360 AKTSETNAKASSETSAESSKTAASASSASSAS 393

RESULT 6
TS1024
related to C2H2 zinc finger transcription factor D-Spl [imported] - Neurospora crassa
N:Alternate names: protein B7F21.50
C:Species: Neurospora crassa
C:Date: 21-Jul-2000 #sequence_revision 21-Jul-2000 #text_change 21-Jul-2000
R:Schulte, U.; Aign, V.; Hoheisel, J.; Brandt, P.; Fartmann, B.; Holland, R.; Nyakatu
submitted to the Protein Sequence Database, July 2000
A:Reference number: Z25286
A:Accession: TS1024
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-770 <SCH>
A:Cross-references: EMBL:AL389901; GSPDB:GN00116; NCSP:B7F21.50
A:Experimental source: BAC clone B7F21; strain OR74A
C:Genetics:
A:Gene: NCSP:B7F21.50
A:Map position: 6
A:Introns: 117/1

Query Match 5.2%; Score 148.5; DB 2; Length 770;
Best Local Similarity 19.9%; Pred. No. 0.25; Mismatches 151; Indels 227; Gaps 25;
Matches 109; Conservative 62

QY 5 NLGHNPNVNSIPPAPPLPSQTDG---AGGRCQLINST---GPLGSR-----ALF 48
DB 219 NSGAIPKFLPSSVPKASAIPTKDKSDGSGG---LNGAQLGGIIGGATALLIIVIAAF 274
QY 49 TPVR---NSMADSGDNRAVDVGLPVNPMRLAASEITLNDGFEVLHDHGPLDLNROIGSS 106
DB 275 LIIRLKRIVESAMESKKGSTSGYHSKASKTSQAQMEQSGRFLHVRAPSDTDNASADPL 334
QY 107 VRFVETQEDGKHI-----AVGO--RNGVETSVVLS-----DQ 136
DB 335 MFMSETNTPGDHTTNASSLAGTPQCAHGVGDGSGIGRHRGSGSDTTTYMASHPGAGNNN 394
QY 137 EYARLOSIDP-----EGKD-----KFVFTGGRGAGHAMVTVASDITEARQRI 181
DB 395 NGSELASDPNRYFDGASPPPLPSSSHTTTYYAGCHNGHGM-----RESV-- 442
QY 182 LLEPKGTGSKGSGKGVGLRESNSGAENTTETQTSTSSLSRDPKLLWALGT----- 237
DB 443 -----DSQSTGLGYHYSTTRNQHQHHRNQSNASELSAD-----GSEITH 483
QY 238 -VATGLIG-----LAATGIV-----QALATPEPDS-----PTTDPD 269
DB 484 GVASPLVGGSSHARGASGTSYRYTHTHSHSHSGLDVDPDSSGMFVELPATTTTTT 543
QY 270 AAASAT-----ETATRDQLTKEAFQNPDN-----QKVNIDELGNAIPSGVLKDDVVA 316

[illegible]

A>Note: the authors translated the codon GTT for residue 1103 as Tyr
A>Note: part of this sequence, including the amino end of the mature protein, were detected
C:Genetics:
A:genome: plasmid
C:Superfamily: lactocepin; subtilisin homology
C:Keywords: hydrolase; serine proteinase; transmembrane protein
F:1-33/Domain: signal sequence #status predicted <SIG>
F:34-187/Domain: propeptide #status predicted <PRO>
F:188-1962/Product: serine proteinase, cell-envelope-associated #status predicted <MAT>
F:208-634/Domain: subtilisin homology #status atypical <SBT>
F:1938-1955/Domain: transmembrane #status predicted <TM>

Query Match 5.0%; Score 141.5; DB 2; Length 1962;
Best Local Similarity 22.8%; Pred. No. 2.3;
Matches 114; Conservative 61; Mismatches 211; Indels 113; Gaps 20;
Qy 36 INSTGPGSRALFTVPR-----NSMADSGDNRASDPGLPVPNMR-LAASFTLND 85
Db 1452 INSGKP-GHMAIDQPVKLEGGKLVAVTDSDDNTTKNITVYEPKTLAAPTVPFS- 1509
Qy 86 GFEVLHDHGLDNLNRQIGSSVFRVETQDGHIAVGORNGVETSVLSQDEYARLQSID 145
Db 1510 -----TTEPAQTTLTNAANA-----TGETVQYSADGGKTYQDVP-----AAGVTIT 1551
Qy 146 PGKDKFVTGGGCGAGHAMVTVASDI-----TEARQRIELELLEPKGTGESKAGES 197
Db 1552 ANGTFKXSTLDYGNESPAVDYVVTNIKADDPQAQLQAQKQELNLIASAKTLSASGKYDD 1611
Qy 198 KGVGELRESNGAENTTTQSTSTSSLR-SDPKLWLALGTATGLIGLAATGIVQALAL 256
Db 1612 ATTTALAAATQKAQ-TALDQINASVDSLGTANRDLOTAINLAALPADKKTSLNLQLOS 1670
Qy 257 TPE-----PSPPTTDPDAASATETATROLKEA---FQNPDKQKVNDELGNATPSGVL 310
Db 1671 VKDALGTDLGNOTDP--STGKFTTAAALDLVAQAAGTQTDQLQATLAK-----I 1719
Qy 311 KDDVVAANTEQAKAGE---EAKQAI-----ENNAQAQKYDEQQAQR 351
Db 1720 LDEVLAKLAEGIKAAATPAEVGNAKDAATGKWTADYADTLTSGQASADSKLAHLQALQ 1779
Qy 352 QBELKVSS-----GAGYGLSGALILGGGIGVAVTA-----ALHRKNQP 389
Db 1780 SLKTKVAAVAEAAKTVGKGGTGTGSDKGGGQTPAPAGDTGDKDGEGSQPSGGNIP 1839
Qy 390 VQQTNTTTTTTTTSGAR-----TVENKPNANTPAQGNVD---TPGSEDVWESRRSM- 438
Db 1840 TKPATTTSTTTTDDTRNGQLTSGTSDKGGGQTPAPAGDTGDKDGEGSQPSGGNIP 1899
Qy 439 ----ASTSSTFTDSSIGG 453
Db 1900 TNPATTTSTTTDDTRNG 1918

RESULT 14
A40315
maternal effect protein (staufen) - fruit fly (Drosophila melanogaster)
C:Species: Drosophila melanogaster
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C:Accession: A40315
R:St. Johnston, D.; Beuchle, D.; Nusslein-Volhard, C.
Cell 66, 51-63, 1991
A:Title: staufen, a gene required to localize maternal RNAs in the Drosophila egg.
A:Reference number: A40315; MUID:91300552
A:Accession: A40315
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1026 <STJ>
A:Cross-references: GB:M69111; NID:g158505; PIDN:AAA73062.1; PID:g158506
C:Genetics:
A:Gene: FlyBase:stau
A:Cross-references: FlyBase:FBgn0003520
C:Superfamily: maternal effect protein; double-stranded RNA-binding repeat homology

F:308-379/Domain: double-stranded RNA-binding repeat homology <DSR1>
F:575-646/Domain: double-stranded RNA-binding repeat homology <DSR2>
F:708-782/Domain: double-stranded RNA-binding repeat homology <DSR3>

Query Match 4.9%; Score 140.5; DB 1; Length 1026;
Best Local Similarity 20.0%; Pred. No. 1.1;
Matches 97; Conservative 74; Mismatches 207; Indels 107; Gaps 18;
Qy 2 PIG-NLGNPNVNSIPPAPLPPLPSQTDAGGAGQLINSTGPL-----GSRALFT----- 49
Db 501 PLGAHVHGP--NGFPSPVTPPPSKITLFFVKQKQFVGRTLLQQAQKHAARALQVLTKQ 558
Qy 50 -----PVRNSMADSGDNRASDPGLPVPNMR-LAASEITLNDGFEVLHDHGLDNLNRQ 102
Db 559 AISASEALEDSM-DEGDKS-----PISQVHEIGIKRNMVHFVKVLRREGPAHMKNF 611
Qy 103 IGSSVFRVETQDGHIAVGORNGVETSVLSQDEYARLQSIDPEGKDKF-----VFTGGR 158
Db 612 TACIVGSIVTEGEGNGKVKKRAAEKMLV-----ELQKLPLPLTPKQTPKRIKVKTPGK 667
Qy 159 GGAGHAMVTVAS-----DITEARQRIELELLEPKGTGESKG 193
Db 668 SGAAAREGSVVGTDGTMTQGTGKPKERRKRLNPPKDLIDMDADNPITKLILQLOTRK--- 724
Qy 194 AGESKGVGELRESNG-----AENTTETQSTSTSSLRSDPKLWLALGTATGLIGLAAT 248
Db 725 --ERPIFELIAKKNETARREFVNEVSASGRTARGTGNKK-LAKRNAQAQLFEL--- 778
Qy 249 GIVQALALTPEPDSPTTDPDAASATETATRDQLTKEAFQNPDKQKVNDELGNATPSG 308
Db 779 --LEAVQVTP-----TNETQSSECCTSATMSAVTAPAVEATAEGKPMVATPVGPWP 830
Qy 309 VL-----KDDVV---ANIE-----EQAKAAGEAKQOAIENNAQAQKYDE 346
Db 831 ILILRQNKPAKRDQIVIVKSNVESKEEANKAVAAEENSNNSANGSDSSSSSGDS 890
Qy 347 QOAKQEELKYSSGAGYGLSGALILGGGIGVAVTAALHRKNQPVQETTTTTTTTTTSAR 406
Db 891 QATEAASESALNTSGNTSGVSSNSNVGANTDGNNAESK--NNTESSSNSTNTQSA 948
Qy 407 TVENK 411
Db 949 GVHMK 953

RESULT 15
T24583
hypothetical protein T06D8.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T24583
R:Palmer, S.
submitted to the EMBL Data Library, April 1995
A:Reference number: Z19909
A:Accession: T24583
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1829 <WIL>
A:Cross-references: EMBL:Z49130; PIDN:CAA88964.1; GSPDB:GN00020; CESP:T06D8.1
C:Genetics:
A:Gene: CESP:T06D8.1
A:Map position: 2
A:Introns: 1391/3; 1432/3; 1470/3; 1505/1; 1520/1; 1616/1; 1644/1; 1687/3; 1742/1

Query Match 4.9%; Score 140.5; DB 2; Length 1829;
Best Local Similarity 22.3%; Pred. No. 2.4;
Matches 106; Conservative 64; Mismatches 191; Indels 115; Gaps 21;
Qy 12 VNNSIPPAPLPSPQTDAGGAGQLINSTGGLSRALFTPVNSMADSGDNRASDPVGLPV 71

Thu Sep 27 15:30:01 2001

Db 180 VTASSAATTTAGTEASGEETTTSAVTEGSGEETTVAVVES---SGEPASSSTTSIP- 235
Qy 72 NPMRLAASEITLNDGFVLHDHGPLDTLNRQIGSSVFRVETQEDGKHIAVGQRNGVETSV 131
Db 236 -----TELSKND--QVTEASGE-ETI-----TAAATEASEETTTSAVTEGSGEDTTV 279
Qy 132 VL-----SDQEYARLQSIDPE-GDKRFVFGGRGGAGHAMVTVASDITEARQRILELLEP 185
Db 280 VAVVELSGEQPSSSTSIPTELSKDDQV-----TEASGEETTTAA--TEASE-----ET 327
Qy 186 KGTGESKGAGESKGVGELRESNCAENTTETQTSTSTS---SLRSDPKLWLALGTVATGL 242
Db 328 TTSVAVTEGSGEETTVAVVLESSG-----EPPASSSTSIPTELSKDDQVTEASGEETT-- 379
Qy 243 IGLAATGIVQALALTPDPDSPTTTDDPAAASATETATRDQLTKEAFONPDNOKVNIDELG 302
Db 380 -----TAAATEASEETTT-----SAVTEGSGEDTTVVAVVLESSGEQPAS---SS 420
Qy 303 NAIPSGVLKDDV--ANTEEQAAGAEAKQAATENNAQAQKYDEQQAQRQELKYSSG 360
Db 421 TSIPTELSKDDQVTEASGEETTTAAATEASEETTS-----AVTEG 461
Qy 361 AGYGLSGALIL-----GGGIGVAVTAALHRKNQPV-----EQTTTTTTT-----TTTTSAR 406
Db 462 SGEDTTVVAVVLESSGEQPASSSTSIPTELSKDDQVTEASGEETTTAAATEASEETTTSAV 521
Qy 407 T-----VENKPANTPAGNVDTPGSEDTWESRRESSMASTSTFFDTS 450
Db 522 TEGSGEETTVAVVLESSGEQPASSSTSIPTELSKDDKVTEASGEETTTAAATDASS 577

Search completed: September 27, 2001, 14:21:47
Job time: 83 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 27, 2001, 14:22:50 ; Search time 14.89 Seconds
(without alignments)
1286.019 Million cell updates/sec

Title: US-09-189-415A-4
Perfect score: 2851
Sequence: 1 MPIGNLGHNPVNNISPPAP.....SNSAVNTSNPPAPGSHRFV 559

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	146.5	5.1	1120	1	STFR_ECOLI
2	142	5.0	563	1	MUCS_BOVIN
3	140.5	4.9	1026	1	STAU_DROME
4	136.5	4.8	461	1	US45_LACLC
5	133.5	4.7	796	1	Y58A_CAEEL
6	132.5	4.6	1258	1	ICEN_ERWHE
7	130.5	4.6	1034	1	ICEN_PANAN
8	130	4.6	1077	1	HLES_DROME
9	129.5	4.5	2411	1	DAB_DROME
10	129	4.5	784	1	SP4_HUMAN
11	129	4.5	2090	1	HFCL_MESAU
12	128.5	4.5	1902	1	P3P_LACLC
13	127.5	4.5	3726	1	TRX_DROME
14	127	4.5	1337	1	DEXT_STRDO
15	126	4.4	997	1	BIR1_SCHPO
16	125.5	4.4	1322	1	ICEA_PANAN
17	125	4.4	1694	1	IGA0_HAEIN
18	125	4.4	1702	1	IGA2_HAEIN
19	125	4.4	1772	1	MSP1_PLAYO
20	125	4.4	1802	1	HKR1_YEAST
21	125	4.4	1902	1	P1P_LACLC
22	124	4.3	1902	1	P2P_LACLC
23	123.5	4.3	2483	1	PCX_DROME
24	123	4.3	1849	1	IGA4_HAEIN
25	122.5	4.3	2842	1	APC_RAT
26	122	4.3	2003	1	YDBA_ECOLI
27	122	4.3	2035	1	HFCL_HUMAN
28	121.5	4.3	1079	1	IF2P_SCHPO
29	121.5	4.3	1189	1	YJH6_YEAST
30	121	4.2	1150	1	APMU_PIG
31	121	4.2	1306	1	MSB2_YEAST
32	120.5	4.2	746	1	7UP2_DROME
33	120	4.2	797	1	VGLX_HSVB

34	119	4.2	484	1	P60_LISMO
35	119	4.2	1048	1	P100_RCMVA
36	119	4.2	1581	1	PPRB_HUMAN
37	119	4.2	1603	1	PSC_DROME
38	118.5	4.2	516	1	P54_ENTFC
39	118	4.1	662	1	MUC1_XENLA
40	118	4.1	1781	1	AKAC_HUMAN
41	117.5	4.1	523	1	P60_LISSE
42	117.5	4.1	1025	1	SIAP_CAUCR
43	117	4.1	567	1	CHI3_CANAL
44	117	4.1	681	1	VGP_MABVP
45	117	4.1	1036	1	NIT2_NEUCR

ALIGNMENTS

RESULT 1
STFR_ECOLI
ID STFR_ECOLI STANDARD; PRT: 1120 AA.
AC P76072: P77560;
DT 01-OCT-2000 (Rel. 40, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DE 01-OCT-2000 (Rel. 40, Last annotation update)
DE SIDE TAIL FIBER PROTEIN HOMOLOG FROM LAMBDOID PHOPHAGE RAC.
GN STFR OR B1372.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-K12;
RX MEDLINE=97251357; PubMed=9097039;
RA Alba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K., Itoh T.,
RA Kasai H., Kashimoto K., Kimura S., Kitakawa M., Kitagawa M.,
RA Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,
RA Nakade S., Nakamura Y., Nashimoto H., Nishio Y., Oshima T., Saito N.,
RA Sampei G., Seki Y., Sivasubram S., Tagami H., Takeda J.,
RA Takemoto K., Takeuchi Y., Wada C., Yamamoto Y., Horiuchi T.;
RT "A 570-kb DNA sequence of the Escherichia coli K-12 genome
corresponding to the 28.0-40.1 min region on the linkage map.";
RL CORRESPONDING TO THE 28.0-40.1 min region on the linkage map.";
CC -1- SIMILARITY: BELONGS TO THE TAIL FIBER FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AE000234; AAC74454.1; ALT_INIT.
DR EMBL: D90774; BAA14966.1; -.
DR EMBL: D90775; BAA14975.1; -.
DR EcoGene; EG13370; stfr.
DR HSP; P04002; 1WFA.
DR InterPro; IPR000122; -.
DR Hypothetical protein; Fiber protein; Repeat.
SQ SEQUENCE 1120 AA; 113779 MW; 542E59D71EE795B4 CRC64;

	DR	PROSITE; PS01208;	VWFC; 1.
	DR	PROSITE; PS01185;	CTCK_1; 1.
	DR	PROSITE; PS01225;	CTCK_2; 1.
	KW	Repeat; Glycoprotein.	
	FT	DOMAIN	61..158
	FT	REPEAT	61..71
	FT	REPEAT	112..122
	FT	REPEAT	148..158
	FT	DOMAIN	338..404
	FT	DOMAIN	471..555
	FT	SIMILAR	1
	FT	DISULFID	471..518
	FT	DISULFID	485..532
	FT	DISULFID	494..548
	FT	DISULFID	498..550
	FT	DISULFID ?	?
	FT	CARBOHYD.	28..28
	FT	CARBOHYD	394..394
	FT	CARBOHYD	477..477
	SEQ	SEQUENCE	563 AA; 58913 MW; AB326CD7BE5FFCF CRC64;
		Query Match	5.0%; Score 142; DB 1; Length 563;
		Best Local Similarity	20.3%; Pred No. 0.29;
		Matches	75; Conservative 48; Mismatches 169; Indels 78; Gaps
	QY	88 EVLHDHGPLDTLNROIGSSVFRVET-----OEDGKHIAVGORNGVETSVLVS	134
	DB	2 KVLQENSPRHAIIS---GSSHTEATTLIVSNSTSGTGLRPEDNTAVAGGOATGRVT----	53
	QY	135 DQEYARLQSIDPEGDOKFYPTGGRGAGHAMVTVASDIPEARORILELLPDKGTGESKA	194
	DB	54 -----GTKK-VLPGETTVAPGSSNTESTTSLSGESRTRIGRI-----TGATTGT	94
	QY	195 GESKGYGELRESNSGA-ENTTETQTSTSSSLSDPKMLALGTVAATGLIGAATGIIVA	253
	DB	95 SKRSSPGS-KTGNTGALSCTTVAPGSSNTGATTS----LGSETTQGQIKIVTMGVTTG	148
	QY	254 LAITPDPSPTTTDPAAASATETATRDQLTKFAONPNQKNVIDELGNALPSGVLKDD	313
	DB	149 TTIIAPGSSNKATTPTEVRTTTEVRATEFT-----TSRHSSDATSGSIOTGI----	196
	QY	314 VVANIEBQAAGEAEAKQAOIENNAOAQQKYDEQQAQRBELKVSSGAGYGLSGALIGG	373
	DB	197 -----TGTSGETTSPFGFNAAETTFKEH-----VRTTETRLSGTTTRGSRTVIPE	244
	QY	374 GIGVAVTAALHRKNQPVEQTTHTTTTTSARTVENKNPANNPAGNVDTPCSEDTMES	433
	DB	245 SSNTGSTGVGR-----QSSTAVSVGRVTGV--SESSSPGTSKEASETTTTGPGIS	297
	QY	434 RRSMSASTSS	443
	DB	298 SSKNRITTSS	307
		RESULT 3	
	ID	STAUI_DROME	STANDARD; PRT; 1026 AA.
	AC	P25159;	
	DT	01-MAY-1992 (Rel. 22, Created)	
	DT	01-MAY-1992 (Rel. 22, Last sequence update)	
	DT	01-OCT-2000 (Rel. 40, Last annotation update)	
	DE	MATERNAL EFFECT PROTEIN STAUFEN.	
	GN	STAUF.	
	OS	Drosophila melanogaster (fruit fly).	
	OC	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;	
	OC	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;	
	OC	Ephydroidea; Drosophilidae; Drosophila.	
	RX	NCBI_TaxID=7227;	
	RA	[1]	
	FP	SEQUENCE FROM N.A.	
	RP	MEDLINE=91300552; PubMed=1712672;	
	RA	St Johnston D., Beuchle D., Nuesslein-Volhard C.;	

Query Match 4.8%; Score 136.5; DB 1; Length 461;
Best Local Similarity 19.1%; Pred. No. 0.47; Mismatches 163; Indels 14; Gaps 14;
Matches 84; Conservative 69;

QY 181 ELLEPKGTGSGKAGSGKVGELRESNGAETTTQTSTSSLRSDPKLMLALGTAT 240
DB 32 DIAKQATISSAQAQAQAVDSLSQKVSLSQKQOTSTKAQIAKSELKALNAQIAT 91
QY 241 GLIGL-AATGIVQALALTPEDPSTTTDPDA-----AASATEATRDOL-- 283
DB 92 LNESIKERTKLEAQAARSQAQVNSATNMDAVVNSKSLTDVIOKVTAIATVSSANKQILE 151
QY 284 -----KFAQFNDKQVNDI--LGNALPS-----GVLKDDVVANIEEQAKAAG 326
DB 152 QQKEQKELSKQETVKKNVNFVLSQSLDSQAQELTSSQAELK---VATLNYQATIA 208
QY 327 EEAKQQAIAENNAQAQKYDQQAQKQ--ELKYSSGAGYGLSGLILGGIGVAVTAALH 384
DB 209 AQDKQALLDEKAAKAAQAAQAAQAAQAAQAAQAAQAAQAAQAAQAAQAAQAA 268
QY 385 RKNQPVQTTTTTTTTTARTVENKPAANTPAQGNVDTPGSEDTHMESRRSSMASTTS 444
DB 269 SSSQAPQVSTQSTQSTQSTQSTQSTQSTQSTQSTQSTQSTQSTQSTQSTQST 324
QY 445 FFDSSIGGPGCRIRMLKHKRMRRCRLILIRLFRINGIQISVVYSTIQHPPTDTN 504
DB 325 GTSTGNTGG-----TTTG 337
QY 505 GARL-----LGNPSAG-----IQTYARLALSG-----GLRHDG-- 534
DB 338 GSGINSPIGNPYAGGCTDYVMOYFAAQGIYIRNPGNGGQWASGPAQGVLLHYVGA 397
QY 535 -GLTGGNSA--VNTSNP 550
DB 398 PGVIASSFADFVGANSP 416

RESULT 5
YS8A_CAEEL STANDARD; PRT; 796 AA.
AC Q09625;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE HYPOTHETICAL 84.3 KDA PROTEIN ZK945.10 IN CHROMOSOME II.
GN ZK945.10.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderiinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Wilkinson-Sproat J.;
RL Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Z48582; CAA88469.1;
DR EMBL; Z48544; CAA88469.1; JOINED.
DR EMBL; Z48544; CAA88444.1; -.
DR EMBL; Z48582; CAA88444.1; JOINED.
DR WormPep; ZK945.10; C601732.
KW Hypothetical protein; Transmembrane.
FT TRANSMEM 11 30 POTENTIAL.
FT DOMAIN 273 546 SER/THR-RICH.
FT DOMAIN 636 732 SER/THR-RICH.

SQ SEQUENCE 796 AA; 84306 MW; 76DC5B03E6357A6A CRC64;
Query Match 4.7%; Score 133.5; DB 1; Length 796;
Best Local Similarity 19.7%; Pred. No. 1.4;
Matches 82; Conservative 76; Mismatches 176; Indels 83; Gaps 14;

QY 68 GLPVNPMRLAASEITLND-----GFEVLHDHPLDLNRLQIGSSVFRVETQEDGKHI 119
DB 137 GLFLNSTWITLNEVNDDEISIAVEAKYEVCYDDG----IDRCDSGLMW---LQVGGNEM 189
QY 120 A-VGQNGVETSVLSQYARLQSDPEGKDKFVETGGRGGAGHAMVTVASDITEARQR 178
DB 190 ALIGYREKCESEI--NEETARMCKRPYRSEK-----STAISDSQGV 230
QY 179 ILELLPKPTGSGKAGSGKVGELRESNGAETTTQTSTSSLRSDPKLMLALGTAT 238
DB 231 YDGVQVKGVRAKQFSMTSGSPTLRMRKRDAGNCTDYTIESTSTSTTTTPTTSTTV 290
QY 239 -ATGLIGLAATGIVQALALTPEDPSTTT-----DPDAASATETATRDQLTKEAFQ 289
DB 291 TSTTTVPTSTSTVTAMSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTST 350
QY 290 NPDQKVNIDELGNAIPSGVLKD--DVANIERQAKAAG--EEAKQQAIAENNAQAQKYDE 346
DB 351 SPSS-----TTLSTSTPTTTTPEITSLSLPDAICSYLDETTTFTTTTMTLTSTTTE 405
QY 347 QQAKRQBELKVSAGYGLSGALILGGIGVAVTAALHRRKNQVPEOTTTTTTTTTT--S 404
DB 406 EPSTSTTTTEVTS-----TSSVTVTTEPTTTTSTSTSTSTSTSTSTSTSTST 445
QY 405 ARTVENKPA-----NNTPAQGNVDTPGSEDTHMESRRSSMASTTSSTFTSSIG 452
DB 446 TSTVTTSPSTVSTVSTVSTVSTVSTVSTVSTVSTVSTVSTVSTVSTVSTVSTV 502

RESULT 6
ICEN_ERWHE STANDARD; PRT; 1258 AA.
ID ICEN_ERWHE
AC P16239;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE ICE NUCLEATION PROTEIN.
GN ICEE.
OS Erwinia herbicola.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Erwinia.
OX NCBI_TaxID=557;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-M1;
RX MEDLINE=90152370; PubMed=2515997;
RA Warren G.J., Corotto L.V.;
RT "The consensus sequence of ice nucleation proteins from Erwinia
RT herbicola, Pseudomonas fluorescens and Pseudomonas syringae.";
RT Gene 85:239-242(1989).
RL CC
CC -! FUNCTION: ICE NUCLEATION PROTEINS ENABLE BACTERIA TO NUCLEATE
CC CRYSTALLIZATION IN SUPERCOOLED WATER.
CC -! SUBCELLULAR LOCATION: OUTER MEMBRANE.
CC -! DOMAIN: CONTAINS 126 IMPERFECT REPEATS OF A CONSENSUS OCTAPEPTIDE
CC A-G-Y-G-S-T-X-T. FURTHER ON A 16-RESIDUE AND A REGIONAL 48-RESIDUE
CC PERIODICITY IS SUPERIMPOSED.
CC -! MISCELLANEOUS: A STRUCTURAL MODEL IS SUGGESTED IN WHICH THE ICE
CC NUCLEATION PROTEIN DISPLAYS A SYMMETRY RELATED TO THAT OF ICE.
CC -! SIMILARITY: BELONGS TO THE BACTERIAL ICE NUCLEATION PROTEIN
CC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>;
or send an email to license@isb-sib.ch).

CC -----
DR EMBL; M26382; AAA24823.1; -.
DR PIR; JQ0188; JQ0188.
DR HSSP; P06620; LINA.
DR InterPro; IPR000258; -.
DR Pfam; PF00818; Ice_nucleation; 65.
DR PRINTS; PR00327; ICENUCLEATN.
DR PROSITE; PS00314; ICE_NUCLEATION; 45.
KW Ice nucleation; Repeat; Outer membrane.
FT DOMAIN 162 1217 OCTAPEPTIDE PERIODICITY.
SQ SEQUENCE 1258 AA; 125084 MW; 590E8A130077FBD4 CRC64;

Query Match 4.6%; Score 132.5; DB 1; Length 1258;
Best Local Similarity 20.4%; Pred. No. 2.8; Mismatches 226; Indels 141; Gaps 20;
Matches 110; Conservative 62;
QY 26 TDGAGGRGLI---NSTGTLGSRALFTVPRNSMADSGDNRAVDVPLPVPNPMRLAASEI- 81
DB 183 TETAGDSSTLGIAGYGTGTAGADSTLVAGYGTQTAGE-ESSQAGYGTGTGMKGSDLT 241
QY 82 -----TLNDGFEVLHDHGPLDT-----LNRTGSSVFRVETQEDGKHIAVGORN--- 125
DB 242 AGYGTGTAGDSSSLIAGYGTGTAGDSSSLTAGYGT---QTAQKGSDLTAGYGTGT 297
QY 126 -GVETSVVLSDOEYARLQSIDPE-----GKDKFVTGGRGAG----- 162
DB 298 AGADSSLIAG---YGTGTAGEESTQTAGYGTGTQAQKGSDLTAGYGTGTAGDSSSLIA 354
QY 163 -----HAMVTVASDITEARQRIELLEPKGTGSKGAGESKVGCELRESNSGAEN 212
DB 355 GYGTGTAGDSSSLTAGYGTGTQAQKGSDLTAGYGTGTAGADSSSLIAGYGTGTAGEES 414
QY 213 T-TETQTSTSSLSRDPKMLALGTAVATGLIG-----LAATGIVQALALTPEDSPPT- 265
DB 415 TOTAGYGTGTQAQKGS-----LTAGYGTGTAGDSSSLIAGYGTGT---TAGEDSSSLTA 466
QY 266 -----TDPDAASATETATRDQLTKEAF-----QNPDKNQK 295
DB 467 GYGTGTQAQKGSDLTAGYGTGTAGESSSLIAGYGTGTAGYGTGTLAGYGTGTQAQNES 526
QY 296 VNIDELGNAIPSGVLKDDVAVANIEQAQAEEAKQQAQAIENNAQAQKQYDEQ 355
DB 527 DLITGYGTSTAGA-----NSSLIAGYGTGTASVNSVLTAGYGTGTAREQSVLTTCYGSTS 576
QY 356 KVSAGYGLSGAL-----ILGGGIGVAVTAALH-----RKNQPVETTTTTT 399
DB 577 T-----AGYGTGTAGDSSSLIAGYGTGTASVHSSLTAGYGTGTAREQSVLTTCYGSTS 632
QY 400 TTTTSARTVENKPNANTPAQGNVDTPGSEDIMESRRSS-----MASTSTFTDSSIGG 453
DB 633 TAGADSSLIAGYGTGTAGYNSILTAGYGTGTQAQKGSDLTAGYGTGTAGADSSLIAG 691

RESULT 7

ICEN_PANAN
ID ICEN_PANAN STANDARD; PRT; 1034 AA.
AC Q47879;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE ICE NUCLEATION PROTEIN INAU.
GN INAU.
OS Pantoea ananas (Erwinia uredovora).
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Pantoea.
OX NCBI_TaxID=553;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=KUIN-3;
RX MEDLINE=94264407; PubMed=7764866;

RA Michigami Y., Watabe S., Abe K., Obata H., Arai S.;
RT "Cloning and sequencing of an ice nucleation active gene of Erwinia
uredoovora";
RL Biosci. Biotechnol. Biochem. 58:762-764(1994).
CC -!- FUNCTION: ICE NUCLEATION PROTEINS ENABLE BACTERIA TO NUCLEATE
CRYSTALLIZATION IN SUPERCOOLED WATER.
CC -!- SUBCELLULAR LOCATION: OUTER MEMBRANE.
CC -!- DOMAIN: CONTAINS IMPERFECT REPEATS OF A CONSENSUS OCTAPEPTIDE
A-G-Y-G-S-T-X-T; FURTHER ON A 16-RESIDUE AND A REGIONAL 48-RESIDUE
PERIODICITY IS SUPERIMPOSED.
CC -!- MISCELLANEOUS: A STRUCTURAL MODEL IS SUGGESTED IN WHICH THE ICE
NUCLEATION PROTEIN DISPLAYS A SYMMETRY RELATED TO THAT OF ICE.
CC -!- SIMILARITY: BELONGS TO THE BACTERIAL ICE NUCLEATION PROTEIN
FAMILY.

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

CC -----
DR EMBL; D14992; BAA03636.1; -.
DR HSSP; P06620; LINA.
DR InterPro; IPR000258; -.
DR Pfam; PF00818; Ice_nucleation; 51.
DR PRINTS; PR00327; ICENUCLEATN.
DR PROSITE; PS00314; ICE_NUCLEATION; 34.
KW Ice nucleation; Repeat; Outer membrane.
FT DOMAIN 162 993 OCTAPEPTIDE PERIODICITY.
SQ SEQUENCE 1034 AA; 103378 MW; FA22523D333EADD CRC64;

Query Match 4.6%; Score 130.5; DB 1; Length 1034;
Best Local Similarity 20.9%; Pred. No. 2.9; Mismatches 229; Indels 139; Gaps 23;
Matches 114; Conservative 64;
QY 26 TDGAGGRGLI---NSTGTLGSRALFTVPRNSMADSGDNRAVDVPLPVPNPMRLAASEI- 81
DB 183 TETAGDSSTLGIAGYGTGTAGDSTLVAGYGTQTAGE-ESSQAGYGTGTGMKGSDLT 241
QY 82 -----TLNDGFEVLHDHGPLDT-----LNRTGSSVFRVETQEDGKHIAVGORN--- 125
DB 242 AGYGTGTAGDSSSLIAGYGTGTAGDSSSLTAGYGT---QTAQKGSDLTAGYGTGT 297
QY 126 -GVETSVVLSDOEYARLQSIDPE-----GKDKFVTGGRGAG----- 162
DB 298 AGADSSLIAG---YGTGTAGEESTQTAGYGTGTQAQKGSDLTAGYGTGTAGDSSSLIA 354
QY 163 -----HAMVTVASDITEARQRIELLEPKGTGSKGAGESKVGCELRESNSGAEN 212
DB 355 GYGTGTAGDSSSLTAGYGTGTQAQKGSDLTAGYGTGTAGADSSSLIAGYGTGTAGEES 414
QY 213 T-TETQTSTSSLSRDPKMLALGTAVATGLIG-----LAATGIVQALALTPEDSPPT- 265
DB 415 TOTAGYGTGTQAQKGS-----LTAGYGTGTAGDSSSLIAGYGTGT---TAGEDSSSLTA 466
QY 266 -----TDPDAASATETATRDQLTKEAF-----QNPDKNQK 295
DB 467 GYGTGTQAQKGSDLTAGYGTGTAGESSSLIAGYGTGTAGYGTGTLAGYGTGTQAQNES 526
QY 296 VNIDELGNAIPSGVLKDDVAVANIEQAQA-----AGEEAKQQAQAIENNAQAQKQYDEQ 348
DB 527 DLITGYGTSTAGA-----NSSLIAGYGTGTASVNSVLTAGYGTGTAREGSDLTAGYGTGT 585
QY 349 AKRQELKVSSG-----AGYGLSGALILGGGIGVAVTAALHR-----KNQPVETTTTTT 397
DB 586 AQENSDLTTCYGTSTAGYDSS-----LIAGYGTGTAGVHSLTAGYGTGTQAQERSDLT 641
QY 398 T-----TTTTTART--VENKPNANTPAQGNVDTPGSEDIMESRRSS-----MASTSTFTD 447
DB 642 TGYGTSTAGADSSSLIAGYGTGTAGYNSILTAGYGTGTQAQKGSDLTAGYGTSTAGYE 701

Best Local Similarity 18.9%; Pred. No. 3.2;
Matches 145; Conservative 66; Mismatches 244; Indels 312; Gaps 29;

QY 448 TSSIG 453
Db 702 SSIAG 707

```
RESULT 8
HLES_DROME STANDARD; PRT; 1077 AA.
AC Q02308;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 01-FEB-1994 (Rel. 28, Last annotation update)
DE HAIRLESS PROTEIN.
GN H.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92387549; PubMed=1516831;
RA Bang A.G., Posakony J.W.;
RT "The Drosophila gene Hairless encodes a novel basic protein that
RT controls alternative cell fates in adult sensory organ development.";
RL Genes Dev. 6:1752-1769(1992).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=93041287; PubMed=1419850;
RA Maier D., Stumm G., Kuhn K., Prells A.;
RT "Hairless, a Drosophila gene involved in neural development, encodes
RT a novel, serine rich protein.";
RL Mech. Dev. 38:143-156(1992).
CC -!- FUNCTION: IS A POTENT ANTAGONIST OF NEUROGENIC GENE ACTIVITY
CC DURING SENSORY ORGAN DEVELOPMENT. THE EXPRESSION OF DISTINCT CELL
CC FATES BY THE TRICHOGEN (SHAFT) / TORMOGEN (SOCKET) SISTER CELL
CC PAIR DEPENDS ON THE LEVEL OF H ACTIVITY. A CERTAIN THRESHOLD LEVEL
CC OF H ACTIVITY IS REQUIRED, BELOW WHICH BOTH SISTER CELLS ADOPT THE
CC TORMOGEN FATE.
CC -!- SUBCELLULAR LOCATION: NUCLEAR (PROBABLE).
CC -!- TISSUE SPECIFICITY: OVARY, EMBRYOS, LARVAL AND PUPAL IMAGINAL
CC DISCS.
CC -!- SIMILARITY: CONTAINS A "PRD MOTIF".
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M95192; AAA28607.1; ALT_INIT.
DR EMBL; X67239; CAA47664.1; -.
DR HSSP; P04002; 1WFA.
DR FlyBase; FBgn0001169; H.
KW Developmental protein; Nuclear protein; DNA-binding.
FT DOMAIN 115 123 THR-RICH.
FT DOMAIN 642 648 POLY-SER.
FT DOMAIN 879 891 POLY-ALA.
FT DOMAIN 937 946 POLY-ALA.
FT DOMAIN 964 974 ALA-RICH.
FT DOMAIN 979 1008 HIS/PRO-RICH (PRD MOTIF).
FT CONFLICT 151 151 S -> A (IN REF. 2).
FT CONFLICT 702 703 OH -> LL (IN REF. 2).
FT CONFLICT 891 891 A -> R (IN REF. 2).
FT CONFLICT 964 967 RAVA -> RLLP (IN REF. 2).
FT CONFLICT 974 974 MISSING (IN REF. 2).
SQ SEQUENCE 1077 AA; 111039 MW; A94BFIA27579E2F1 CRC64;
```

Query Match 4.6%; Score 130; DB 1; Length 1077;

```
QY 7 GHNPVNNNSIPPAPP-----LPSQDAGGGRGQLINSTGPLGS--RALF 48
Db 92 GLNGSSSTSYPLPPPLPANLSRTTPTTTTPSSSSSTASNGFLPHAKTPKSSSMAAS 151
QY 49 TPNRNSMADSGDNRAS-DVPG-----LPVNPMLAAASEIT-----LNDGF 87
Db 152 AAVAASVVCATASKPTIDVLGGVLDYSSLGAATGSLTATVVAAGATAKIGKNSGG 211
QY 88 EYLDHGGPLDTL-NRQIGSSVFVETQEDGKHIAVGQRNGVETSVVLSDDQYARLQSIDP 146
Db 212 SFDMGRTPLSTHGNNSWGYGGRLOFFKDKGKFL-----ELARSKDGDK 255
QY 147 EGKDKFVETGGR--GGAGHAMVTVASDITEARQRILELLEPKGTGSGKAGESKGVGELR 204
Db 256 SGWVSVTRKTRPPSAATSATVPTSAVTTA-----YPRN-----290
QY 205 ENSGAENTETQTSTSTSSLSRSDP-----KLWLALGTVATGLIGLA 246
Db 291 -----ENSTLSFSDDNSSIOSPPQRDPQWKQPRRGISKELSLFFHPRNSTLGRA 344
QY 247 ATGIVQALALTPEPDSPTTDDP-----AAASAFE-----TATRDOL 283
Db 345 A--LRTAARKRRPHEPLTSEDQQIFATAIKAENGDDDTLKAAEAAVEIENVAVDTT 402
QY 284 TKE-AFQNPDP-----NOKVN 297
Db 403 TNEIKIEKPDITKGEDDAERLEKEPKKAVSDDSEKASPGQOQVEPQKDETVDVEMKKN 462
QY 298 IDE-----LGNAPISGVLDKDWANIEQAKAAGEAKQQAIAENNAQAQKYDEO 347
Db 463 TSEDEEPMELPRITNAV-NGDLNGDLKASIGPKPKPKAKLSSI-----IQKLDSV 516
QY 348 QAKRQBELKVSSGAGYGLSGALILGGGIGVAVTAALHRKNQPVQETTT-----395
Db 517 PARLEQMSKTSAVIASTTTSSDRIGGGLSHALT---HKVSPSSATAAGRLVEYHTQHSV 573
QY 396 -----TTTTTTTTSARTVENKPNANTPAQ 419
Db 574 PRKRILREFEKVLEDNGCVNNGSGAGGAGGKRSRAKGTSTSSPAGKASPNNLAPPQ 633
QY 420 GNVD-TPGSE-----DTMESRRSSM-----ASTSTFTTSTSGGCRIRMLM 461
Db 634 GKPSFGSSSSSTSPATLSTQPTRLNSSYSIHSLGGSGSGSSSSSSSGKKC-----687
QY 462 LKHRCMIRRCRLILIRLFRWGIQISVYVSTIOHPPRDTDNGARLLGNPSAGIQSTYA 521
Db 688 -----GDHPAAIISNVHHPQHS-----WYQFSS---SSYP 714
QY 522 RLALSGGLRHDMDGLTGG-----SNSAVNTSNPPAPGS-----HRF 558
Db 715 RALLTSPKSPDVSGSNGGKGKSPSHTGTRKRSPPYSAGSPVDYGHSE 761
```

RESULT 9

```
DAB_DROME STANDARD; PRT; 2411 AA.
ID AC P98081;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE DISABLED PROTEIN.
GN DAB.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93194063; PubMed=7680635;
```



```

DR PRINTS: PR000048; ZINC_FINGER.
DR PROSITE; PS000028; ZINC_FINGER_C2H2.1; 3.
DR PROSITE; PS00157; ZINC_FINGER_C2H2.2; 3.
KW Transcription regulation; Activator; Zinc-finger; Metal-binding;
KW DNA-binding; Nuclear protein; Repeat.
FT DOMAIN 7 11 POLY-GLU.
FT DOMAIN 12 19 POLY-ALA.
FT DOMAIN 122 130 POLY-SER.
FT DOMAIN 165 188 POLY-SER.
FT DOMAIN 647 729 ZINC_FINGERS.
FT ZN_FING 647 729 C2H2-TYPE.
FT ZN_FING 677 701 C2H2-TYPE.
FT ZN_FING 707 729 C2H2-TYPE.
FT CONFLICT 197 197 K -> Q (IN REF. 2).
FT CONFLICT 379 380 HA -> QP (IN REF. 2).
FT CONFLICT 386 386 Q -> A (IN REF. 2).
FT SEQUENCE 784 AA; 82025 MW; 3C4EAE28CB81FB CRC64;

Query Match 4.5%; Score 129; DB 1; Length 784;
Best Local Similarity 20.0%; Pred. No. 2.5;
Matches 112; Conservative 71; Mismatches 234; Indels 144; Gaps 21;

QY 3 IGNLGHNPVNNNS---IPAPPL-----PSQ-----TDAGGRGQLINSTGPIG-SRAL 47
DB 58 IGTPEGNAQTGQQIITDPSQGLVQLQNQPQQLVLTQTLAGNAWQLVASTPPASKENNV 117
QY 48 FTPVRNMAISGDNRAISDVFGPLPNPRLAASEITLNDGFEVLHDHGPDLTLNRQIGSV 107
DB 118 SQAPASSSSSSSSNNNGS-----ASPTKSGNSSTPGQFQVQLVQNPSSGVQVIVPOL 171
QY 108 FRVETQEDGKHIAVGQNGVETSIVLSDQBYARLQSDPEGKDFEFT-GRGGAGHAMV 166
DB 172 QTVEGQI-----QINPTSSSLQDLQGIKLS---AGNNQAULTAARNTASGNILA 221
QY 167 -----TVASDITEARQILLELPKGT-----GESK 192
DB 222 QNLANTVPVQIRPGVSIPLQTLPTGTAQGVVTLPLNIGGVTLALPVINNVAAGSGTG 281
QY 193 GAGESKGVGLRESNGAENTETSTSTSLRSLDKPLALGTAT-----GLIGL 245
DB 282 QVGPAATADSGTNGNLQVSTPTNTTSTASTMPESSTTCTTASTSLTSSDTLVSS 341
QY 246 AATGIVQALALTPEDSTPTTDPDAASATETATR--DQLTKEAFONPDNOKVNIDELG- 302
DB 342 ADTG---QYASTSASSERTIEESQTPAATESAQSSSLHANGMNQDQNSLQOVQI 398
QY 303 -----NAIP-----SG-----VLKDDVWVNIIEQA----- 322
DB 399 VQPILOIQIQPOQOQIIQAIIPPOSFOLQSGQTIQTQQPQLQNVQLQVNPQVLIRA 458
QY 323 ---KAAGBEAKQQALENNAQAOKKYDEQQAQRQELKYSSGAGYGLSALLIGGGIGVAV 379
DB 459 PLTSPGQISQMTQVQVQNIQSLQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQV 515
QY 380 TAAHRKNQPVQEQTTTTTTTTTTTTSARIVENKPAANTPAQGNVDTPGS-----EDTME 432
DB 516 AGA-----PITLNTAQLASVFNLTQSVANLGAAGVQVQVQVQVQVQVQVQVQVQV 569
QY 433 SRRSSMASTSSFTFTSSIGG 453
DB 570 VQOATIAPV-----TVAVGG 584

RESULT 11
HFC1_MESAU
ID HFC1_MESAU STANDARD; PRT; 2090 AA.
AC P51611;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE HOST CELL FACTOR C1 (HCF) (VP16 ACCESSORY PROTEIN) (HFC1) (VCAF)
DE (CFF).

GN Mesocricetus auratus (Golden hamster).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC Mesocricetus.
OX NCBI_taxid=10036;
RN [1]
RP SEQUENCE FROM N.A.
RA Goto H., Nishitani H., Umene K.I., Nakabeppu Y., Nishimoto T.;
RL Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: UPON LYTIC INFECTION OF PERMISSIVE CELLS, THE HSV
CC TRANSACTIVATOR PROTEIN VP16 ASSOCIATES WITH HCF. BINDING TO HCF
CC ACTIVATES VP16 FOR ASSOCIATION WITH THE OCTAMER MOTIF-BINDING
CC PROTEIN OCT-1, TO FORM A MULTIPROTEIN-DNA COMPLEX RESPONSIBLE FOR
CC ACTIVATING TRANSCRIPTION OF THE HSV IMMEDIATE EARLY GENES (BY
CC SIMILARITY).
CC -!- SUBUNIT: COMPOSED PREDOMINANTLY OF SIX POLYPEPTIDES RANGING FROM
CC 110 TO 150 KDA AND A MINOR 300 KDA POLYPEPTIDE. THE MAJORITY OF N-
CC AND C-TERMINAL CLEAVAGE PRODUCTS REMAIN TIGHTLY, ALBEIT
CC NONCOVALENTLY, ASSOCIATED (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: NUCLEAR.
CC -!- DOMAIN: THE HCF REPEAT IS A HIGHLY SPECIFIC PROTEOLYTIC CLEAVAGE
CC SIGNAL (BY SIMILARITY).
CC -!- PTM: EXISTS AS A FAMILY OF POLYPEPTIDES THAT ARISE THROUGH
CC CLEAVAGE AT ONE OR MORE CENTRALLY LOCATED SITES. CLEAVAGE OCCURS
CC AT A DEFINED SITE, PPCE/THET, WITHIN THE HCF REPEAT (BY
CC SIMILARITY).
CC -!- SIMILARITY: CONTAINS 5 KELCH REPEATS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; DA5419; BAA08258.1;
CC HSSP; P02751; 1FNA.
CC InterPro; IPR001777;
CC Pfam; PF00041; fn3; 2.
CC Nuclear protein; Repeat.
CC REPEAT 44 89
CC FT REPEAT 93 140 KELCH 1.
CC REPEAT 148 194 KELCH 2.
CC REPEAT 217 265 KELCH 3.
CC REPEAT 266 313 KELCH 4.
CC REPEAT 313 348 KELCH 5.
CC FT DOMAIN 1010 1448 8 X 26 AA APPROXIMATE REPEATS.
CC REPEAT 1010 1035 HCF REPEAT 1.
CC REPEAT 1072 1097 HCF REPEAT 2.
CC REPEAT 1101 1126 HCF REPEAT 3.
CC REPEAT 1157 1182 HCF REPEAT 4 (DEGENERATE).
CC REPEAT 1295 1320 HCF REPEAT 5.
CC REPEAT 1323 1348 HCF REPEAT 6.
CC REPEAT 1358 1383 HCF REPEAT 7 (DEGENERATE).
CC REPEAT 1423 1448 HCF REPEAT 8.
CC SEQUENCE 2090 AA; 214942 MW; E495E8B1F2385E17 CRC64;

Query Match 4.5%; Score 129; DB 1; Length 2090;
Best Local Similarity 19.3%; Pred. No. 8.6;
Matches 121; Conservative 71; Mismatches 224; Indels 224; Gaps 27;

QY 3 IGNLGHNPVNNNSIPPAPPLPSQTDGAGRG---QLINSTGPGLSRALFTPVNRNMAISG 59
DB 677 ISNIG---KVMVSVQTKPVQISAVTQASTGPVTQIIQTKGPLPAGTIL-----KLVTSA 728
QY 60 DNRAISDV-----PGLPVNPMPLAASEI---TLNDGFEVLHDHGPDLTLNRQIGSVFV 110
DB 729 DGKPTTIITTTQASGAGSKPTILGIGSSVSPSTTKPGTTTIKTIPMSAIIQAGAT---- 784
QY 111 ETQEDGKHLAVQGRNVETSVVLSQVEY-----ARLQSIDP-----EGDKKFFV 154

```


Db 785 -----GVTSTPGIKSPITIIITKVTMTSGTAPAKIITAVPKIATGCHGOQVTVVL 835
QY 155 TCGRGAGHAMVTV-----ASDIT-EARORILELLEPKGTGESKGAGSGVGLRES 206
Db 836 KGAPQGAAILRVPMGSRVLPVTVSAVPAVTVLVVKGTTGTTGLTGTGTSTSLA 895
QY 207 NSGAEVTTETSTSTSLRSDPKLWALGAVVATGLIGLAATGIVQALALPEPDSPTT 266
Db 896 GAGAHSTASLATPIT-----LGIAT-----LSQVINPAITV----- 931
QY 267 DPDAASATETATRDQLKEAFQFNDKVINDELGNAPISGVLKDDVVANIEQAKAAG 326
Db 932 ---SAAQTLLTAAGLTPITIMQPSQPTQVTLI--TAPSGV----- 969
QY 327 BEAKQALENNAQAKKYD-----EQAKRQELKVS-----SGAGYGLSGALLGGGIV 377
Db 970 -----EQPVHDLPSVILASPTTQPTAIVTIADSGQGVQPG----- 1007
QY 378 AVTALHRKNQPVETTTTTTTTTTTTTSARTVENKPNANTPAQ----- 419
Db 1008 ---TVTLVCSNPPCETHETGTTNTATT--VVANLGGHPQPTQVQFVCDRQEAASLVTS 1063
QY 420 -----GNVDTPGSEDTMESRSSMASTSTFTDSSIGGPCRIMLMKLRHRCMIRCRLL 474
Db 1064 VQQNGNVVRCNSNPPCETHETGTTNTATT--ATSNMAG-----QHGC----- 1104
QY 475 ILIRLFRWIGIISVYTIQHPHRTDNGARLLGNFSAGIQSTYARLALSGGLRHDWG 534
Db 1105 -----SNPPCETHETG-----TSTATTAMS-----SMG 1128
QY 535 GLTGCSNAVNTSNPP-----APGS 555
Db 1129 ---TGOQDRHSSNPTVVRITVARGA 1153
RESULT 12
P3P_LACLC
ID P3P_LACLC STANDARD; PRT; 1902 AA.
AC P15292;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE PIII-TYPE PROTEINASE PRECURSOR (EC 3.4.21.96) (LACTOCEPIN) (CELL WALL-
DE ASSOCIATED SERINE PROTEINASE).
GN PRTP.
OS Lactococcus lactis (subsp. cremoris) (Streptococcus cremoris).
OG Plasmid.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Lactococcus.
OX NCBI_TaxID=1359;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 188-197.
RC STRAIN=SK11;
RX MEDLINE=89304035; PubMed=2760036;
RA Vos P., Simons G., Siezen R.J., de Vos W.M.;
RT "Primary structure and organization of the gene for a procaryotic,
RT cell envelope-located serine proteinase."
RL J. Biol. Chem. 264:13579-13585(1989).
CC -1- FUNCTION: PROTEASE WHICH BREAKS DOWN MILK PROTEINS DURING THE
CC GROWTH OF THE BACTERIA ON MILK.
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. CELL WALL.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8; ALSO KNOWN AS THE
CC SUBTILASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DB EMBL; J04962; AAA03533.1; ALT_SEQ.

PIR: A32634; A32634.
DR HSP: P00782; 2SEF.
DR MEROPS: S08.019; .
DR InterPro: IPR000209; .
DR Pfam: PF00746; Gram_pos_anchor; 1.
DR Pfam: PF00082; peptidase_S8; 3.
DR PRINTS: PR00723; SUBTILISIN.
DR PROSITE: PS00136; SUBTILASE_ASP; 1.
DR PROSITE: PS00137; SUBTILASE_HIS; 1.
DR PROSITE: PS00138; SUBTILASE_SER; 1.
DR PROSITE: PS00343; GRAM_POS_ANCHORING; 1.
KW Hydrolase; Serine protease; Cell wall; Zymogen; Signal; Plasmid;
KW Transmembrane.
FT SIGNAL 1 33
FT PROPEP 34 187
FT CHAIN 188 1902
FT DOMAIN 188 1876
FT TRANSMEM 1877 1895
FT DOMAIN 1896 1902
FT ACT_SITE 217 217
FT ACT_SITE 281 281
FT ACT_SITE 620 620
FT DOMAIN 1867 1872
FT SEQUENCE 1902 AA; 200550 MW; 87CECBAA9345F9D3 CRC64;
Query Match 4.5%; Score 128.5; DB 1; Length 1902;
Best Local Similarity 22.7%; Pred. No. 8.1;
Matches 105; Conservative 56; Mismatches 197; Indels 105; Gaps 18;
QY 36 INSTGPLGSRALFTVPR-----NSWADSGDNRAVDVPGLPVNPWR-LAASEITLND 85
Db 1452 INSGKP-GHMAIDQPVKLEGNKLVTAVDSEDTNTTKNITVYEPKTKLAAPTPTS- 1509
QY 86 GFEVLHDHGPDLTLNRQIGSSVRFVETQDGHIAVGORNGVETSVLSLDQYARLQSID 145
Db 1510 -----TTEPACTVTLTANAAA---TGFTVQYSADGGKTYQDVP-----AAGVTIT 1551
QY 146 PEGKDKFVTGGRGAGHAMVTVASDI-----TEARQRIELLEPKGTGESKGAGS 197
Db 1552 ANGTEFKFSTDLGYNESPAVDVVTNIRAKDDPAQLQAQKQELTNLIASAKTILSASGKYDD 1611
QY 198 KGVGELRESNGAENTTETOTSTSSLR-SDPKLWALGTVATGLIGLAATGIVQALAL 256
Db 1612 ATTTALAATQKQAQ-TALDQTNASVDSLTGANRDLQTAINQLAAKLPADKKTSLNLQLOS 1670
QY 257 TPE---PDSPTTTPDAAASATETATRDQLTKEA---FQNPQNKVNIDELGNAPISGVL 310
Db 1671 VKDALGTDLGNQTDTP--STGKTTFTAAALDDLVQAQAGTQTDQLQATLAK-----I 1719
QY 311 KDDVVANIEEQAKAAGE---EAKQQA-----ENNAQAKKYDEQQAAR 351
Db 1720 LDEVLAKEAGIEKAATPAEVGNAKDAATGKTWYADIADTLTSGQASADASDKLAHLQALQ 1779
QY 352 QEELKVSS-----GAGYGLSGALLGGGIGVAVTA-----ALHRKNOP 389
Db 1780 SUKTVAAAVEAAKTVGKGDGTTGSDGGGGQGTAPADPGDTGKDKGDEGSPSSGGNIP 1839
QY 390 VEQTITTTTTTTTTTSARTVENKPNANTPAQGNVDTPGSEDTME 432
Db 1840 TKPATTTTITDITDTR-----NGQLTSGKGLPKTGETTE 1875
RESULT 13
TRX_DROME
ID TRX_DROME STANDARD; PRT; 3726 AA.
AC P20659; Q27255; Q27327;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE TRITHORAX PROTEIN.

Transcription regulation: zinc-finger; Metal-binding; DNA-binding;
Nuclear protein; developmental protein; Activator;
Alternative splicing.

Accession	Protein Name	Length	Score	DB 1	Length	DB 2
QY	QY1266	1347	1266	1347	1347	1347
QY	QY1348	1393	1348	1393	1393	1393
QY	QY1421	1482	1421	1482	1482	1482
QY	QY1734	1793	1734	1793	1793	1793
QY	QY1794	1844	1794	1844	1844	1844
QY	QY3599	3708	3599	3708	3708	3708
QY	QY512	516	512	516	516	516
QY	QY565	570	565	570	570	570
QY	QY661	664	661	664	664	664
QY	QY905	910	905	910	910	910
QY	QY1576	1582	1576	1582	1582	1582
QY	QY2298	3027	2298	3027	3027	3027
QY	QY3032	3040	3032	3040	3040	3040
QY	QY3181	3184	3181	3184	3184	3184
QY	QY3220	3225	3220	3225	3225	3225
QY	QY1	368	1	368	368	368
QY	QY2025	2025	2025	2025	2025	2025
QY	QY2341	2341	2341	2341	2341	2341
QY	QY2392	2392	2392	2392	2392	2392
QY	QY3726	3726	3726	3726	3726	3726

Query Match 4.5%; Score 127.5; DB 1; Length 3726;
Best Local Similarity 22.0%; Pred. No. 22;
Matches 89; Conservative 55; Mismatches 148; Indels 113; Gaps

Accession	Protein Name	Length	Score	DB 1	Length	DB 2
QY	QY176	176	176	176	176	176
QY	QY866	866	866	866	866	866
QY	QY236	236	236	236	236	236
QY	QY923	923	923	923	923	923
QY	QY295	295	295	295	295	295
QY	QY962	962	962	962	962	962
QY	QY355	355	355	355	355	355
QY	QY1002	1002	1002	1002	1002	1002
QY	QY415	415	415	415	415	415
QY	QY1048	1048	1048	1048	1048	1048
QY	QY474	474	474	474	474	474
QY	QY1086	1086	1086	1086	1086	1086
QY	QY513	513	513	513	513	513
QY	QY1146	1146	1146	1146	1146	1146

Query Match 4.5%; Score 127.5; DB 1; Length 3726;
Best Local Similarity 22.0%; Pred. No. 22;
Matches 89; Conservative 55; Mismatches 148; Indels 113; Gaps

Accession	Protein Name	Length	Score	DB 1	Length	DB 2
QY	QY176	176	176	176	176	176
QY	QY866	866	866	866	866	866
QY	QY236	236	236	236	236	236
QY	QY923	923	923	923	923	923
QY	QY295	295	295	295	295	295
QY	QY962	962	962	962	962	962
QY	QY355	355	355	355	355	355
QY	QY1002	1002	1002	1002	1002	1002
QY	QY415	415	415	415	415	415
QY	QY1048	1048	1048	1048	1048	1048
QY	QY474	474	474	474	474	474
QY	QY1086	1086	1086	1086	1086	1086
QY	QY513	513	513	513	513	513
QY	QY1146	1146	1146	1146	1146	1146

Query Match 4.5%; Score 127.5; DB 1; Length 3726;
Best Local Similarity 22.0%; Pred. No. 22;
Matches 89; Conservative 55

Thu Sep 27 15:30:03 2001

```
DR PROSITE: PS01282; BIR_REPEAT_1; FALSE_NEG.
DR PROSITE: PS50143; BIR_REPEAT_2; 2.
KW Cell Division; Repeat.
FT REPEAT 25 99 BIR 1.
FT REPEAT 120 194 BIR 2.
FT DOMAIN 80 83 POLY-ASP.
FT DOMAIN 312 319 POLY-ASP.
FT DOMAIN 487 490 POLY-SER.
SQ SEQUENCE 997 AA; 112579 MW; 952A6BAFA5C489FA CRC64;

Query Match 4.4%; Score 126; DB 1; Length 997;
Best Local Similarity 19.8%; Pred. No. 5;
Matches 98; Conservative 62; Mismatches 205; Indels 130; Gaps 17;

QY 13 NNSIPPAPPLPSQTDGAGRGQLINSTGLGSRALFTPVNSMADSGDNRASDVPLPVN 72
Db 222 DNSILPVSPTRDSTKSHKTLNFSRKNLNARPLTMSLYTNTSEKDSQPTRAPQSPTK 281
QY 73 PMRLAASEITLNDGFEVLHDHGLDILNRQIGSS-----VFRVETQED----- 115
Db 282 PVLLTAPR-----RKNKSPKSKPAVFKPVKPIFSDDEDDDLTASQ 324
QY 116 -----GKHIAVGQRNGVETSVVLSQDEYARLQSDPEGKDKFVFTGGRGGAGHAMVTV 168
Db 325 PFSKGICNDSMQVAKKNFTETIPLKDEKNELEHLVSPATS--VHTTVSDITGHOSVTD 382
QY 169 ASDITEARQRIELLEPKGTGSGKAGESKGVGELRESNSGA-----ENTTETQTST 222
Db 383 ESD-----EQNNCMSTPPKIEIESIYVSKSKEISSVSVSGVKEQNHTEKQVAIET 438
QY 223 SSL-----RSDPKLWALGTVATGLIGLAATGIQALALTPEPDSPTTTPDAAASATETA 278
Db 439 PEQKQVKEDEHLN-----QGSFIEESTKQPISSKPSSTSPD---MTDAA 481
QY 279 TRDQLTKRAFQNPQNKVNIDELGNATIPSGVLKDDVVANIEEQAKAAGEAKQQAIIENNA 338
Db 482 TGGRVSSSSFRDKILO-----NFSRSTI--DSFSNISK--KRNSEANDENDETNL 530
QY 339 Q---AQKYDEQAKRQELKAVSSGAGYGLSGALLGGIGVAVTAALHRKNQPVQETTT 395
Db 531 KIPPEKRRKFQEVLSQKNILVSS-----TEDSHEPVKYTED 567
QY 396 TTTTNTTTTARTVENKPNANTPA-----QGNVDTP-----GSEDTMES---RRS 436
Db 568 SQTAHVSKFEDLENKSMESQSLQLLSESEDDKPLIDLIPLLAIKRNKDLNLSGVLEKG 627
QY 437 SMATSTSTFTFTSSI 451
Db 628 KSTSTSTKTKFDTISV 642
```

Search completed: September 27, 2001, 14:22:56
Job time: 152 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 27, 2001, 14:27:03 : Search time 38.06 Seconds
(without alignments)
1943.208 Million cell updates/sec

Title: US-09-189-415a-4

Perfect score: 2851

Sequence: 1 MFIGNLGHNPVNNNSIPPAP.....SNSAVNTSNPPAPGSHREV 559

Scoring table: BLOSUM62 /

Gapop 10.0 , Gapext 0.5

Searched: 425026 seqs, 132305027 residues

Total number of hits satisfying chosen parameters: 425026

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL16:*

1: sp_archaea:*

2: sp_bacteria:*

3: sp_fungi:*

4: sp_human:*

5: sp_invertebrate:*

6: sp_mammal:*

7: sp_mhc:*

8: sp_organelle:*

9: sp_phase:*

10: sp_plant:*

11: sp_rodent:*

12: sp_unclassified:*

13: sp_vertebrate:*

14: sp_virus:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Length	DB ID	Description
1	2624.5	92.1	558	2 Q9R396	Q9R396 escherichia
2	2608.5	91.5	558	2 Q85506	Q85506 escherichia
3	1774.5	62.2	538	2 Q47014	Q47014 escherichia
4	1766.5	62.0	538	2 Q47016	Q47016 escherichia
5	1753.5	61.5	538	2 Q85508	Q85508 escherichia
6	1570.5	55.1	551	2 Q88258	Q88258 escherichia
7	1554.5	54.5	552	2 Q9KWH9	Q9KWH9 escherichia
8	1548.5	54.3	550	2 Q52147	Q52147 escherichia
9	1522.5	53.4	547	2 Q9WXX1	Q9WXX1 escherichia
10	1521.5	53.3	547	2 Q9ETI1	Q9ETI1 citrobacter
11	1495.5	52.4	549	2 Q50190	Q50190 escherichia
12	190.0	6.7	1323	5 Q9NHX4	Q9NHX4 drosophila
13	170.0	6.0	1229	5 Q94185	Q94185 caenorhabdi
14	165.5	5.8	2232	5 P91365	P91365 caenorhabdi
15	162.5	5.7	1291	5 Q77261	Q77261 drosophila
16	162.5	5.7	1300	5 Q9W5E0	Q9W5E0 drosophila
17	161.0	5.6	2586	5 Q9VTK8	Q9VTK8 drosophila
18	156.0	5.5	2921	5 Q9N973	Q9N973 leishmania
19	152.5	5.3	1412	5 Q9VKJ1	Q9VKJ1 drosophila

20	149.5	5.2	676	5 Q9VB94	Q9VB94 drosophila
21	148.5	5.2	770	3 Q9P319	Q9P319 neurospora
22	148.0	5.2	1192	5 Q17346	Q17346 caenorhabdi
23	148.0	5.2	1192	5 Q17346	Q17346 caenorhabdi
24	147.5	5.2	3012	5 Q97W45	Q97W45 leishmania
25	145.0	5.1	332	5 Q9V8Q1	Q9V8Q1 drosophila
26	144.5	5.1	3507	5 Q23587	Q23587 caenorhabdi
27	144.0	5.1	1323	6 Q77509	Q77509 bos taurus
28	143.5	5.0	1275	5 Q76602	Q76602 caenorhabdi
29	143.5	5.0	1805	11 Q63661	Q63661 rattus norv
30	143.0	5.0	2678	5 Q9NDS4	Q9NDS4 dictyostell
31	142.5	5.0	1026	5 Q9V8B9	Q9V8B9 drosophila
32	141.5	5.0	1086	5 Q9W3U2	Q9W3U2 drosophila
33	141.5	5.0	1096	5 Q17368	Q17368 caenorhabdi
34	141.5	5.0	1144	5 Q9TVG6	Q9TVG6 caenorhabdi
35	141.0	4.9	909	13 Q9PTK7	Q9PTK7 xenopus lae
36	141.0	4.9	1475	5 Q9GYW7	Q9GYW7 drosophila
37	140.5	4.9	875	5 Q9W468	Q9W468 drosophila
38	140.5	4.9	1829	5 Q22248	Q22248 caenorhabdi
39	140.0	4.9	576	2 Q9KJ13	Q9KJ13 enterococcu
40	139.0	4.9	2468	2 Q9I2M3	Q9I2M3 pseudomonas
41	138.0	4.8	1038	10 Q9LD52	Q9LD52 oryza sativ
42	137.5	4.8	1385	5 Q9W4Z0	Q9W4Z0 drosophila
43	137.5	4.8	4498	5 Q9W223	Q9W223 drosophila
44	137.0	4.8	620	5 Q9VKL0	Q9VKL0 drosophila
45	137.0	4.8	837	5 P90666	P90666 anthocidari

ALIGNMENTS

RESULT 1

Q9R396	Q9R396	PRELIMINARY;	PRT;	558 AA.
AC	Q9R396;			
DT	01-MAY-2000 (TREMBLrel. 13, Created)			
DT	01-MAY-2000 (TREMBLrel. 13, Last sequence update)			
DT	01-MAR-2001 (TREMBLrel. 16, Last annotation update)			
DE	TRANSLOCATED INTIMIN RECEPTOR TIR.			
GN	TIR.			
OS	Escherichia coli.			
OC	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;			
OC	Escherichia.			
OX	NCBI_TaxID=562;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=86/24;			
RA	MEDLINE=99242825; PubMed=10225900;			
RA	Devlinney R., Stein M., Reinscheid D., Abe A., Ruschkowski S.,			
RA	Finlay B.B.;			
RT	"Enterohemorrhagic Escherichia coli O157:H7 produces TIR, which is			
RT	translocated to the host cell membrane but is not tyrosine			
RT	phosphorylated."			
RL	Infect. Immun. 67:2389-2398(1999).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=EDL933;			
RA	MEDLINE=98339885; PubMed=9673266;			
RA	Perna N.T., Mayhew G.F., Postai G., Elliott S., Donnenberg M.S.,			
RA	Kaper J.B., Blattner F.R.;			
RT	"Molecular evolution of a pathogenicity island from enterohemorrhagic			
RT	Escherichia coli O157:H7."			
RL	Infect. Immun. 66:3810-3817(1998).			
DR	EMBL; AF125993; AAD29391.1;			
DR	EMBL; AF071034; AAC31506.1;			
DR	InterPro; IPR003536;			
DR	PRINTS; PR01370; TRNSINTMINR.			
KW	Receptor.			
QW	SEQUENCE 558 AA; 58022 MW; 99C417222D4B4A1 CRC64;			
Query Match	92.1%;	Score	2624.5;	DB 2; Length 558;
Best Local Similarity	94.0%;	Pred. No.	1e-152;	

Matches 528; Conservative 3; Mismatches 24; Indels 7; Gaps 3;	
QY	1 MPIGNLGNPNVNNNSIPPAPPLPSQTDGAGRGQLINSTGPGSRALFTPVNRNSMADSGD 60
DB	1 MPIGNLGNPNVNNNSIPPAPPLPSQTDGAGRGQLINSTGPGSRALFTPVNRNSMADSGD 60
QY	61 NRASDVPLPVPNMRLAASEITLNDGFEVLHDHGPDLTNRQIGSSVFRVETQEDGKHIA 120
DB	61 NRASDVPLPVPNMRLAASEITLNDGFEVLHDHGPDLTNRQIGSSVFRVETQEDGKHIA 120
QY	121 VGORNGVETSVLSDQYARLQSDPEGDKKFVFTGGRGAGHAMVTVASDITEARQRIIL 180
DB	121 VGORNGVETSVLSDQYARLQSDPEGDKKFVFTGGRGAGHAMVTVASDITEARQRIIL 180
QY	181 ELLEPKGTGSKGAGESKGVGELRESNSGAENTTETQSTSTSLRSDPKLWALGTVA 240
DB	181 ELLEPKGTGSKGAGESKGVGELRESNSGAENTTETQSTSTSLRSDPKLWALGTVA 240
QY	241 GLIGLAATGIVQALALTPEDSPSTTTDPDAAASATETATRDQLTKEAFQNPDKVNIDE 300
DB	241 GLIGLAATGIVQALALTPEDSPSTTTDPDAAASATETATRDQLTKEAFQNPDKVNIDE 300
QY	301 LGNAPISGVLLKDDVVANIEEQAKAAGEEAKQQAENNAQAQKYDEQQAQRQBELKVSSG 360
DB	301 LGNAPISGVLLKDDVVANIEEQAKAAGEEAKQQAENNAQAQKYDEQQAQRQBELKVSSG 360
QY	361 AGYGLSALILGGGIGVAVTAALHRKNQPVQETTTTTTTTTTSARTVENKPNANTPAQG 420
DB	361 AGYGLSALILGGGIGVAVTAALHRKNQPVQETTTTTTTTTTSARTVENKPNANTPAQG 420
QY	421 NVDTPGSEDTMESRRSSMASTSTFFDTSISG---GPCRIRMLMKHRCMIRRCRLILI 477
DB	421 NVDTPGSEDTMESRRSSMASTSTFFDTSISG---GPCRIRMLMKHRCMIRRCRLILI 477
QY	478 RLFRWIGIQISVYVSTIQHPPTDNGARLLGNPSAGIQSTYARLALSGGLRHDGMGLT 537
DB	480 QNM---GNTDSVYVSTIQHPPTDNGARLLGNPSAGIQSTYARLALSGGLRHDGMGLT 536
QY	538 GGNSAVNTSNPPAPGSHRFV 559
DB	537 GGNSAVNTSNPPAPGSHRFV 558
RESULT 2	
085506	PRELIMINARY; PRT; 558 AA.
ID	085506;
AC	01-NOV-1998 (TrEMBLrel. 08, Created)
DT	01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT	01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE	TRANSLOCATED INTIMIN RECEPTOR TIR.
GN	TIR.
OS	Escherichia coli.
OC	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC	Escherichia.
OX	NCBI_TaxID=562;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN-95SP2;
RA	Paton A.W., Manning P.A., Woodrow M.C., Paton J.C.;
RT	"The translocated intimin receptors (Tir) of Shiga toxinogenic
RT	Escherichia coli isolates belonging to serogroups O26, O111, and O157
RT	react with sera from patients with hemolytic-uremic syndrome and
RT	exhibit marked sequence heterogeneity."
RL	Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
DR	EMBL; AF070067; AAC69314.1; .
DR	InterPro; IPR003536; .
DR	PRINTS; PR01370; TRNSINTMINR.
SQ	SEQUENCE 558 AA; 58175 MW; CA2CDDAC94527C2E CRC64;
Query Match	91.5%; Score 2608.5; DB 2; Length 558;
Best Local Similarity	93.4%; Pred. No. 9.5e-152;

RX MEDLINE=99003184; PubMed=9784578;
RA Paton A.W., Manning P.A., Woodrow M.C., Paton J.C.;
RT "Translocated intimin receptors (Tir) of Shiga-toxicogenic Escherichia
coli isolates belonging to serogroups O26, O111, and O157 react with
sera from patients with hemolytic-uremic syndrome and exhibit marked
sequence heterogeneity.";
RL Infect. Immun. 66:5580-5586(1998).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=B10;
RA Marches O., Nougayrede J.-P., Boury M., Mainil J., Charlier G.,
RA Boullier S., De Rycke J., Milon A., Oswald E.;
RT "Role of Tir and Intimin in the pathogenesis of rabbit
enteropathogenic Escherichia coli";
RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; U59502; AAC32028.2; -;
DR EMBL; AJ223063; CAAL1065.1; -;
DR EMBL; AF070068; AAC69316.1; -;
DR EMBL; AF132728; AAD27868.1; -;
DR EMBL; AF113597; AAF03080.1; -;
DR InterPro; IPR003536; -;
DR PRINTS; PRO1370; TRNSINTMINR.
KW Hypothetical protein; Receptor.
SQ SEQUENCE 538 AA; 55420 MW; 31D7A8E227B3D06C CRC64;

Query Match 62.2%; Score 1774.5; DB 2; Length 538;
Best Local Similarity 63.2%; Pred. No. 8.5e-101;
Matches 360; Conservative 61; Mismatches 106; Indels 43; Gaps 6;
QY 1 MPIGNLGNPNVNNISIPAPPPLPSQTDGAGG-RGQLINSTGLGSRALFTPVNRNSMADSG 59
Db 1 MPIGNLGNPNVNNISIPAPPPLPSQTDGAGGARNGLNSGPMGSRLLFTPIRNSVADAA 60
QY 60 DNRA5DVPGLPNMRLAASEITLNDGFVLDHGPDLTLNRQIGSSVFRVETQEDGKH 119
Db 61 DSRASDIPGLPNPLFAASEVSLHGALEVLHDKGGLDTLNSAIGSSLFRVETRDGSHV 120
QY 120 AVQORNGVTSVLSQDEVARLOSIDPEKDKFVFTGGRGAGHAMVTVASDITEARQRI 179
Db 121 AIGQKNGLETTVLSQDEFSLSQSLDPEGKRFVFTGGRGAGHAMVTVASDIAEARQRI 180
QY 180 LELLEPKGSGESKAGESKGVGELRESNGAENTTETQSTSTSLRSDPKLWLAGTVA 239
Db 181 IDKLEPKDKETKEPG-----DPNSGEKKIIEHTSTSTSLRADPKLWLSLGTIA 231
QY 240 TGLIGLAATGIQVALALTEPPDSPPTTDPDAASAATETATRDQLTKEAFQNPQNKVND 299
Db 232 AGLGMAATGIAQAVALTPEPDDPITTPDAAANTAAKADQLTKEAFQNPQNKVND 291
QY 300 ELGNAIPSGVLKDDVVVAIEEQAAGEAKQAQOAKKYEQAQKQAEELKVS 359
Db 292 ENGNALPSGELKDDVVVAIEEQAAGEAKQAQOAKKYEQAQKQAEELKVS 351
QY 360 GAGYGLSGALILGGGIGAVTAALHRKNQPVQETTTTTTTTTTSARTVENKPNNTPAQ 419
Db 352 GVGYGISGALILGGGIGAGVTAALHRKNQPAEQTTITRT-----VVDNQPTNNAQA 403
QY 420 GNVDPGSEDTEMRSSMASTSTSTFFDTS-----IGGPCRIRMLKHKRMTR 469
Db 404 GNTDTSGEESPASRRNSNASLNSGSDTSSTGCTVENPYADVMPRNDLSARIEPIYD 463
QY 470 RCELLILRLFWIGIQISVVSTIQHPDRDITDNCARLLGNPSAGIQSTYARLALSGGL 529
Db 464 E-----VAADPNYSVQIHFGSNPSVPTG-RLVGPFGGQIQSTYALLASSGGL 508
QY 530 RHDMMGLTGSGNSAVNTSNPPAPGSHRFV 559
Db 509 RLGMGGLTGSGESAVSTANAAATPPGPARY 538

RESULT 4
Q47016

ID Q47016 PRELIMINARY; PRT; 538 AA.
AC Q47016;
DT 01-NOV-1996 (TRENBLrel. 01, Created)
DT 01-AUG-1999 (TRENBLrel. 11, Last sequence update)
DT 01-MAR-2001 (TRENBLrel. 16, Last annotation update)
DE TRANSLOCATED INTIMIN RECEPTOR.
GN TIR.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=REPEC 83/39;
RA O'Brien R.A.;
RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=REPEC 83/39;
RA Krejany E.O.;
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=RDEC-1;
RA Agin T.S., Boedeker E.C.;
RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE OF 172-538 FROM N.A.
RC STRAIN=REPEC RDEC-1;
RA Krejany E.O.;
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; U59504; AAD19750.1; -;
DR EMBL; AF045568; AAC15683.1; -;
DR EMBL; U59503; AAB02941.1; -;
DR InterPro; IPR003536; -;
DR PRINTS; PRO1370; TRNSINTMINR.
KW Receptor.
SQ SEQUENCE 538 AA; 55410 MW; 3132A969B7B3D06C CRC64;

Query Match 62.0%; Score 1766.5; DB 2; Length 538;
Best Local Similarity 63.0%; Pred. No. 2.6e-100;
Matches 359; Conservative 61; Mismatches 107; Indels 43; Gaps 6;
QY 1 MPIGNLGNPNVNNISIPAPPPLPSQTDGAGG-RGQLINSTGLGSRALFTPVNRNSMADSG 59
Db 1 MPIGNLGNPNVNNISIPAPPPLPSQTDGAGGARNGLNSGPMGSRLLFTPIRNSVADAA 60
QY 60 DNRA5DVPGLPNMRLAASEITLNDGFVLDHGPDLTLNRQIGSSVFRVETQEDGKH 119
Db 61 DSRASDIPGLPNPLFAASEVSLHGALEVLHDKGGLDTLNSAIGSSLFRVETRDGSHV 120
QY 120 AVQORNGVTSVLSQDEVARLOSIDPEKDKFVFTGGRGAGHAMVTVASDITEARQRI 179
Db 121 AIGQKNGLETTVLSQDEFSLSQSLDPEGKRFVFTGGRGAGHAMVTVASDIAEARQRI 180
QY 180 LELLEPKGSGESKAGESKGVGELRESNGAENTTETQSTSTSLRSDPKLWLAGTVA 239
Db 181 IDKLEPKDKETKEPG-----DPNSGEKKIIEHTSTSTSLRADPKLWLSLGTIA 231
QY 240 TGLIGLAATGIQVALALTEPPDSPPTTDPDAASAATETATRDQLTKEAFQNPQNKVND 299
Db 232 AGLGMAATGIAQAVALTPEPDDPITTPDAAANTAAKADQLTKEAFQNPQNKVND 291
QY 300 ELGNAIPSGVLKDDVVVAIEEQAAGEAKQAQOAKKYEQAQKQAEELKVS 359
Db 292 ENGNALPSGELKDDVVVAIEEQAAGEAKQAQOAKKYEQAQKQAEELKVS 351
QY 360 GAGYGLSGALILGGGIGAVTAALHRKNQPVQETTTTTTTTTTSARTVENKPNNTPAQ 419
Db 352 GVGYGISGALILGGGIGAGVTAALHRKNQPAEQTTITRT-----VVDNQPTNNAQA 403
QY 420 GNVDPGSEDTEMRSSMASTSTSTFFDTS-----IGGPCRIRMLKHKRMTR 469

```

404 GNTDTSPEESPARRRNSNASLNGSDTSTGTVENPYADVGMPRNDSLARISEPIYD 463
470 RCRLLILRLFRINGIOISVYSTIOHPRTDNDGARLLGNPSAGIOSTYARLALSGGL 529
464 E-----VAADPNYSVIOHFSGNSPVTG-RLVGTGQGIQSTYALLASSGGL 508
530 RHDMGGLTGGNSAVNTSNPPAPGSHRFV 559
509 RLGMGLTGGGESAVSTANAPTGPAREV 538

RESULT 5
O85508 ID O85508 PRELIMINARY; PRT; 538 AA.
AC O85508;
DT 01-NOV-1998 (TREMBlrel. 08, Created)
DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
DE TRANSLOCATED INTIMIN RECEPTOR TIR.
GN TIR.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=EP8C87A;
RA Paton A.W., Manning P.A., Woodrow M.C., Paton J.C.;
RT "The translocated intimin receptors (Tir) of Shiga toxinogenic
RT Escherichia coli isolates belonging to serogroups O26, O111, and O157
RT react with sera from patients with hemolytic-uremic syndrome and
RT exhibit marked sequence heterogeneity.";
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF070069; AAC69318.1; -.
DR InterPro; IPR003536; -.
DR PRINTS; PR01370; TRNSINTIMINR.
DR SEQUENCE 538 AA; 55602 MW; 447052A0E3214D6D CRC64;
```

```

Query Match 61.5%; Score 1753.5; DB 2; Length 538;
Best Local Similarity 62.6%; Pred. No. 1.6e-99;
Matches 357; Conservative 61; Mismatches 109; Indels 43; Gaps 6;

Qy 1 MPTGNLGNPNVNNSTPPAPPLPSQTDGAGG-RGOLLNSTGPLGSRALFTPVNRNSMADSG 59
Db 1 MPTGNLGNPNVNNSTPPAPPLPSQTDGAGGAGRNQLINSNGPWSRLFTPLRNSVADAA 60
Qy 60 DNRASDVPGLPNMRLAASEITLNDGFVLDHGDPLDPLNRQIGSSVPRVETQEDGKH 119
Db 61 DSRASDIPGLPTNPLRFAASEVSLHGALVLDHKGGLDTLNSAIGSSLPFRVETRDGSHV 120
Qy 120 AVQORNGVTSVVLSDQEARLOSIDPEGKDFVFTGGRGGAGHAMVTVASDITEARORI 179
Db 121 AIGOKNGLETVYVLSDOEFSSQSLDPEGKNFVFTGGRGGAGHAMVTVASDIAEARORI 180
Qy 180 LELLEPKGTGSKAGSGVGLRESNSGAENTTTQTSTSLRSDPKLWLAIGTVA 239
Db 181 IDKLEPKGTGSKAGSGVGLRESNSGAENTTTQTSTSLRSDPKLWLAIGTVA 231
Qy 240 TGLIGLAATGIVQALATPEPDSPTTTPDPAASATETATROLTKEAQNPNQKVNID 299
Db 232 AGLTGAATGIAQVATPEPDDPTTTPDPAASATETATROLTKEAQNPNQKVNID 291
Qy 300 ELGNAIPSGVLKDDVVANIEEQAAGEAKQAQAIENNAQAQKYDEQAQKQELKVS 359
Db 292 ENGNAIPSGELIDVVVAQIAEQAKAQEQARQAEIESQAQKYDEQAQKQELKVS 351
Qy 360 GAGYGLSGALILGGGIGVAVTAALHRKNQVPEQTITTTTTTTTSARTVENKPNATPAQ 419
Db 352 GVGYGISGALILGGGIGVAVTAALHRKNQVPEQTITTTTTTTTSARTVENKPNATPAQ 403
Qy 420 GNVDTPGSEDTMESRRSSMASTSTSTPDTSS-----IGGPCRIRMLMLKHCIRMIR 469
```

```

404 GNTDTSPEESPARRRNSNASLNGSDTSTGTVENPYADVGMPRNDSLARISEPIYD 463
470 RCRLLILRLFRINGIOISVYSTIOHPRTDNDGARLLGNPSAGIOSTYARLALSGGL 529
464 E-----VAADPNYSVIOHFSGNSPVTG-RLVGTGQGIQSTYALLASSGGL 508
530 RHDMGGLTGGNSAVNTSNPPAPGSHRFV 559
509 RLGMGLTGGGESAVSTANAPTGPAREV 538

RESULT 6
O68258 ID O68258 PRELIMINARY; PRT; 551 AA.
AC O68258;
DT 01-AUG-1998 (TREMBlrel. 07, Created)
DT 01-AUG-1998 (TREMBlrel. 07, Last sequence update)
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
DE TRANSLOCATED INTIMIN RECEPTOR.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98187918; PubMed=9529069;
RA Voss E., Paton A.W., Manning P.A., Paton J.C.;
RT "Molecular analysis of Shiga toxinogenic Escherichia coli O111:H-
RT proteins which react with sera from patients with hemolytic-uremic
RT syndrome.";
RL Infect. Immun. 66:1467-1472(1998).
DR EMBL; AF025311; AAC69249.1; -.
DR InterPro; IPR003536; -.
DR PRINTS; PR01370; TRNSINTIMINR.
DR SEQUENCE 551 AA; 56975 MW; 6EC95F76EF0F44CC CRC64;
```

```

Query Match 55.1%; Score 1570; DB 2; Length 551;
Best Local Similarity 59.5%; Pred. No. 2.7e-88;
Matches 339; Conservative 67; Mismatches 134; Indels 30; Gaps 11;

Qy 1 MPTGNLGNPNVNNSTPPAPPLPSQTDGA--GGRGQLNSTGPLGSRALFTPVNRNSMADS 58
Db 1 MPTGNLGNPNVNNSTPPAPPLPSQTDGASRGAGQLNSTGALGSRLLFSPLRNSIADS 60
Qy 59 GDNRASDVPGLPNMRL--AASEITLNDGFVLDHGDPLDPLNRQIGSSVPRVETQEDG 116
Db 61 VDSR--DIPGLPVHPSRLATATSEICLLGGFEVLHDKGDLTLNKGICASAFRIEQSDG 118
Qy 117 KHIAVGQRNGVTSVVLSDQEARLOSIDPEGKDFVFTGGRGGAGHAMVTVASDITEAR 176
Db 119 SYAAIGERNGVEVSVLNSQELQSLQADIEDKGRFVFTGGRGGGHSWVTPASDIAEAR 178
Qy 177 ORILELLEPKGTGSKAG---ESKGVGELRESNSGAENT--TETQTSTSTSLRSDPKL 231
Db 179 AKILAKLDPNHHGSGOARNVDTSRVGVG---SASGMDSVVSETRTSTSTASSVRSDPKF 234
Qy 232 WLALGTAVATGLIGLAATGIVQALATPEPDSPTTTPDPAASATETATROLTKEAFONP 291
Db 235 WYSIGAIAGLAGLAATGITQALATPEPDDPTTTPDPAASAEASATRDQLTQAFKNP 294
Qy 292 DNOKYNIDELGNAIPSGVLKDDVVANIEEQAAGEAKQAQAIENNAQAQKYDEQAQAKR 351
Db 295 ENOKVSIIDEIGNSIFSGELKDDVVAKIEEQAAGEAARQAQAVESNAQAQRYDYQYARR 354
Qy 352 QBELKVSSGAGYGLSGALILGGGIGVAVTAALHRKNQVPEQTITTTTTTTTSARTVENK 411
Db 355 QEELELSSGIGYGLSGALILGGGIGVAVTAALHRKNQVPEQTITTTTTTTTSARTVENK 408
Qy 412 PANNTPAQGNVDTPGSEDTMESRRSSMASTSTSTFF-DTSSIGGPCRIRMLMLKHCIRMIR 470
Db 409 TGGNTPAQCGGTDAIRAEDTSLNRDRSQRSSTASTHWSDTSSAVVNPYAEVGEARNSSPARQ 468
```


Qy	471	CRLLILRLPRWIGIQISVVYSTIQHPPRD	TTDNGARLLCNFSAGIQSTYARLA	LSGGL	529
Db	469	AEEHIYDE-----VAADPNYSYIQFNSGNQVGTG	RLMGT	PGQIQSTYAILTNN	521
Qy	530	RHDWGGITGGNSAVNTSNPPAPGSHREV			559
Db	522	RLGNGGLTGSGGSAVNTANAAPTPGGRFV			551

```

RESULT      7
Q9KWH9      PRELIMINARY;      PRRT;      552 AA.
ID          Q9KWH9;
AC          AC
DT          01-OCT-2000 (TEMBLrel. 15, Created)
DT          01-OCT-2000 (TEMBLrel. 15, Last sequence update)
DT          01-MAR-2001 (TEMBLrel. 16, Last annotation update)
DE          TIR.
GN          TIR.
OS          Escherichia coli.
OC          Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC          Escherichia.
OX          NCBI_TaxID=562;
RN          [1]
RP          SEQUENCE FROM N.A.
RC          STRAIN=HK01;
RA          Abe A., Nagano H.;
RT          "Analyses of type III secreted proteins and Tir in enteropathogenic
RT          Escherichia coli O157:H45.";
RL          Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
DR          EMBL; AB036053; BAA96815.1; -.
DR          InterPro; IPR003536; -.
DR          PRINTS; PR01370; TRNSINTMINR.
SQ          SEQUENCE      552 AA; 57005 MW;      A8D79EE22EE50A4B CRC64;

```

```

Db      461  - - - - - RNSLSTRQEEHIDEVAADPVYSVIQFNFRNAPVTG-RUMGSPGQGIQ 508
Qy      518  STYARLALSGLRDMGGLTGSGNSAVNT--SNNPPAPGSHRFV 559
      |||||  |||  |||||  |||||  :  ||  |||
Db      509  STYALLANSAGRLGMLGTGSGESAVNTANANAAPTGGPVREV 552
      |||||  |||  |||||  |||||  :  ||  |||

RESULT  8
ID52147 PRELIMINARY; PRT; 550 AA.
AC O52147;
DT 01-JUN-1998 (T-EMBLrel. 06, Created)
DT 01-JUN-1998 (T-EMBLrel. 06, Last sequence update)
DT 01-MAR-2001 (T-EMBLrel. 16, Last annotation update)
DE TRANSLOCATED INTIMIN RECEPTOR.
GN TIR.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RA Elliott S.J., Wainwright L.A., McDaniel T.K., Jarvis K.G., Deng Y.K.,
RA Lai L.C., McNamara B.P., Donnenberg M.S., Kaper J.B.;
RL Mol. Microbiol. 0:0-0(1998).
DR EMBL; AF022236; AAC38390.1; -.
DR InterPro; IPR003536; -.
DR PRINTS; PR01370; TRNSINTMINR.
DR SEQUENCE 550 AA; 56509 MW; 19DD08A9BE9251CB CRC64;
SD

```

Qy	1	MPTGNLGHNPVNNNSITPPAPPLPSOTDGA--GGROLNSTGPGLSRALFTPVRRSMADS	58
Db	1	MPIGNLNHNINISNLLIPTPAPPPLPSTQDGATRGNGSSLSISTSGLSRULLFSPLRSSIVDT	60
Qy	59	GDNRASDVGLPNVMRLAASEITLNDGFEVLHDHGFLDLNRQTIGSSVFVRVETQEDCKH	118
Db	61	VDSK--DVPGLPEHLPRUFATSETCLHGGEVLHDKGFLDLTNOKTIGSSVFVEQQPDGTH	118
Qy	119	IAGVQRNGVETSIVLSDQEVARLOSIDPEGKDKEVFTGTGRCGAGHAMVTVASDIPEAROR	178
Db	119	AALGVKDGVEVSVTLSNSELSQSLSLDTSGETGFYFTGTGCGSGHAMVTVASDISQAREK	178
Qy	179	ILELLEPKGTG-----ESKGAGESKVGBELRESNGAENTTETQTSTSTSLRSDP	229
Db	179	IIAKLPDPNHGRQPDKIDTRSVCGVSASGMGD-----GV--VSETHSTTTSSVRSRDP	230
Qy	230	KMLAGTWATGLIGLAATCIVQALALTPEPDSPTTTPDPAASAATETATRQLTKAEAFQ	289
Db	231	KFWVSGVALAAGLAGLAAGTGIQVAALVTAPDDPTTTDPDEAANAFAATKDLTKAEAFQ	290
Qy	290	NPDNQKNVIDELNAITPSGYLKDDVVANIEEQAKAGEBAKQAIEENNAQAOKKYDEBQA	349
Db	291	NPDNQKNVIDELNAITPSGELKDDVVAIQADQAKVAGEAQARQAVESNAQAQRHHDDQA	350
Qy	350	KROEELKVSSGAGYGILSGALLIGGGIGVATTAALHRKNOPVEQTTTTTTTTTSARTVE	409
Db	351	KROEELDLSGGIYGILSUALIVGGGICAGVTAMLHRNPTEQTIAETHS-----VIQ	403
Qy	410	NKPANNTPAOCNVDTTGESDTEMRESMSMASTSST-FFDTS-SIGGPCRIMLMKHRCM	467
Db	404	QQTGGNTRAOGGADTGVENASLTRRDSQAASVASTQWSDTSGDWNP-----YAEGWM	456
Qy	468	IRCRLLILTLRFINGIQISVYVSTIOHPPRDTTDNGARLLGNPSNGTOSTVARIALSG	527
Db	457	SRRNPSSLAPEPIYDRVADPPNYSVTIOHFSGNNPVTV-RLVGSPGGIGCTVALLASSG	515


```
QY 397 TTTTTSARTVENKPNANTPAQGNVDTPGSEDTMESRRSSMASTSTFFDFTSSIGGPCR 456
Db 1685 TCTMGSTSGTGVSTISSESTASASSQTGVTMTGSSSTGVSSTSS-----ASSGQP-- 1737
QY 457 IRMLMLKHKRMIRRCRLILIRLFRWIGQIS-----VYSTIOHPPRPDRTDNGA 506
Db 1738 -----QMSTSQSSSAGSVVSSSTASPAASSTAPSST 1768
QY 507 RLLGNPSAG-IQSTYARLALSGLRHDMGG-LTGGSNAVNNTSN 549
Db 1769 GTMSSTSGTGVSTMSQSSTAASTTSHGTGVTVLGSSS---TSSN 1810

RESULT 15
O77261 PRELIMINARY; PRT: 1291 AA.
AC O77261;
DT 01-NOV-1998 (TREMBlrel. 08, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
DE EG:115C2.10 PROTEIN.
GN EG:115C2.10 OR CGI3363.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RA Salles C., Valenti P., Darlamitsou A., Henderson N., Campbell L.,
RA Glover D.;
RT "Sequencing the distal X chromosome of Drosophila melanogaster.";
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Benos P.;
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AL031581; CAA20894.2; -
DR FlyBase: FBgn0025639; EG:115C2.10.
DR InterPro: IPR001214; -
DR Pfam: PF00856; SET; 1.
DR SMART: SM00317; SET; 1.
SQ SEQUENCE 1291 AA; 137366 MW; 021DD566196196BA CRC64;

Query Match 5.7%; Score 162.5; DB 5; Length 1291;
Best Local Similarity 20.6%; Pred. NO. 0.088;
Matches 99; Conservative 61; Mismatches 189; Indels 131; Gaps 18;

QY 5 NLGHPNPVNNSTPPAPLPSPOTDG---AGRGQLINSTGLGSRALFTPVNSMADSGDN 61
Db 438 NTSNSNSTNDSTGPSETSTNGLVASGGAG-----GATGAAMLPTP---SQOSTGK 488
QY 62 RAS-----DVPGLPVPMLRA-----ASEITLNDGFEVLHDHGPLDNLNQ 102
Db 489 EATAAVSLLEKKLPNVVSPPLTKMKELQKGMKYDAEMIMANAAYQQQHH----- 539
QY 103 IGVSVFRVETQEDGHIAVGQRNGVETSVVLSQDEYARLQSDPEGKQKFPVFTGGRGAG 162
Db 540 --QHFFHHHHHHHHHHHGHASTGAETAATAVOQMAAMQKPG-----VGGTGAAG 588
QY 163 HAMVTVASDITEARQILLELLEPKETGESKAGESKGVGELRESNAGNTTETOTSTST 222
Db 589 NAGATTVSSVA-----AGAGSEVNGRSTSLRKSMTVNS-----TSSSI 627
QY 223 SSLRSDPKLWLALGTATGLIGLAATGIQALALATPEPDSPTTTPDAAASATETATRDQ 282
Db 628 STASADEVI-----APWAASISLPKAPVVLMPCKPQAQMAIALHQ 670
QY 283 LKTEAFQNPNDQKVNDELGNIPSGVLKDDVVANIEQAKAAGEEAKQO-----A 333
Db 671 SQOQLRRSERQKELTD-----GESSD-----TSSEQOKK-----EQKQDQHLQPKMFS 716
```

```
QY 334 IENNAOAKKYDEQOAKROELKVSSGAGYGLSGALILGGIGVAVTAALHRKNQPVET 393
Db 717 LAEEPQPEKS-EKQOQOQKRVTRNSAGRVGL-----VARLATAHNN-----IA 760
QY 394 TTTTTSARTV---ENKPNANTPAQGNVDTPGSEDTMESRRSSMASTSTFFDFTSS 450
Db 761 TTTNSSSSSNKATTITNCNNHNSNNSSRINHNSNLSRLSVKSRKPAPSEASSIPSTSS 820

Search completed: September 27, 2001, 14:27:07
Job time: 318 sec
```

THIS PAGE BLANK (USPTO)